INTERNATIONAL REVIEW OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY 2003 Volume 18

Edited by
Cary L. Cooper
and
Ivan T. Robertson



International Review of Industrial and Organizational Psychology 2003 Volume 18

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

International Review of Industrial and Organizational Psychology 2003 Volume 18

Edited by

Cary L. Cooper
and
Ivan T. Robertson

University of Manchester
Institute of Science & Technology, UMIST, UK



Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data
International review of industrial and organizational psychology.
—1986—Chichester; New York; Wiley, c1986–
v.: ill.; 24cm.
Annual.
ISSN 0886-1528 = International review of industrial and organizational psychology

1. Psychology, Industrial—Periodicals. 2. Personnel management—Periodicals.
[DNLM: 1. Organization and Administration—periodicals. 2. Psychology, Industrial—periodicals. W1IN832UJ
HF5548.7.157 158.7′05—dc 19 86-643874
AACR 2 MARC-S

Library of Congress [8709]

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0-470-84703-4

Project management by Originator, Gt Yarmouth, Norfolk (typeset in 10/12pt Plantin)
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wilts
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

CONTENTS

ABOUT THE EDITORS

Cary L. Cooper *Manchester School of Management, University of*
Ivan T. Robertson *Manchester Institute of Science and Technology,*
PO Box 88, Manchester, M60 1QD, UK.

Cary L. Cooper received his BS and MBA degrees from the University of California, Los Angeles, his PhD from the University of Leeds, UK, and holds honorary doctorates from Heriot-Watt University, Aston University and Wolverhampton University. He is currently BUPA Professor of Organizational Psychology and Health. Professor Cooper was founding President of the *British Academy of Management* and is a Fellow of the British Psychological Society, Royal Society of Arts, Royal Society of Medicine and Royal Society of Health. He is also founding editor of the *Journal of Organizational Behavior* and co-editor of *Stress and Health*, serves on the editorial board of a number of other scholarly journals, and is the author of over 80 books and 400 journal articles. In 2001, he was honoured by the Queen with a CBE.

Ivan T. Robertson is Professor of Work and Organizational Psychology in the Manchester School of Management, UMIST. He is a Fellow of the British Academy of Management, the British Psychological Society, and a Chartered Psychologist. Professor Robertson's career includes several years experience working as an applied psychologist on a wide range of projects for a variety of different organizations. With Professor Cooper he founded Robertson Cooper Ltd (www.robertsoncooper.com), a business psychology firm which offers consultancy advice and products to clients. Professor Robertson's research and teaching interests focus on individual differences and organizational factors related to human performance. His other publications include 30 books and over 150 scientific articles and conference papers.

CONTRIBUTORS

Helen Baron	*82 Evershot Road, London N4 3BU, UK*
André Büssing	*Technical University of München, Lothstr. 17, D-80335 München, GERMANY*
Ronald J. Burke	*School of Business, York University, Toronto, CANADA*
Peter Y. Chen	*Department of Psychology, Colorado State University, Fort Collins, CO 80523, USA*
Sarah DeArmond	*Department of Psychology, Colorado State University, Fort Collins, CO 80523, USA*
Chris Elshaw	*QinetiQ Ltd: Centre for Human Sciences, A50 Building, Cody Technology Park, Ively Road, Farnborough, Hants, GU14 0LX, UK*
Michael M. Harris	*College of Business Administration, 8001 Natural Bridge Road, University of Missouri, St Louis, MO 63121, USA*
Britta Herbig	*Technical University of Muenchen, Lothstr. 17, D-80335 Muenchen, GERMANY*
Erik Hölzl	*Department of Psychology, University of Vienna, Universitaetsstr. 7, A-1010 Vienna, AUSTRIA*
Erich Kirchler	*Department of Psychology, University of Vienna, Universitaetsstr. 7, A-1010 Vienna, AUSTRIA*
Autumn D. Krauss	*Department of Psychology, Colorado State University, Font Collins, CO 80523, USA*
Suzan Lewis	*Department of Psychology and Speech Pathology, Manchester Metropolitan University, Hathersage Road, Manchester M13 0JA, UK*

Filip Lievens — *Department of Personnel Management, Work and Organizational Psychology, University of Ghent, Henri Dunantlaan 2, 9000 Ghent, BELGIUM*

Tamsin Martin — *SHL Group plc, The Pavillion, 1 Atwell Place, Thames Ditton, Surrey, KT7 0NE, UK*

Lynley H. W. McMillan — *Department of Psychology, University of Waikato, Private Bag 3105, Hamilton, NEW ZEALAND*

Bill Moorcroft — *Sleep and Dreams Laboratory, Luther College, Iowa, USA*

Michael P. O'Driscoll — *Department of Psychology, University of Waikato, Private Bag 3105, Hamilton, NEW ZEALAND*

Ashley Proud — *QinetiQ Ltd: Centre for Human Sciences, A50 Building, Cody Technology Park, Ively Road, Farnborough, Hants, GU14 0LX, UK*

Kirsty Weston — *QinetiQ Ltd: Centre for Human Sciences, A50 Building, Cody Technology Park, Ively Road, Farnborough, Hants, GU14 0LX, UK*

EDITORIAL FOREWORD

The 2003 volume of the *International Review of Industrial and Organizational Psychology* continues with our established tradition of obtaining contributions from several different countries. This edition includes chapters from Germany, Belgium, New Zealand, Austria, Canada, the USA, and the UK. The presence of contributions from such a diverse range of countries indicates the international nature of our discipline. One of the purposes of the international review is to enable scholars from different countries to become aware of material that they might not normally see. We hope that this issue will be particularly helpful in that respect.

Specific issues covered in this volume reflect the growth and complexity of the I/O psychology field. A range of topics from very contemporary issues to well-established topics. The chapter by Lievens and Harris on 'web-based recruiting and testing' and the chapter by Kirchler and Hölzl on 'economic psychology' focus on contemporary topics that we have never dealt with before in the review. Other chapters, such as the review of ethnic differences and cognitive ability by Baron, Martin, Proud, Weston, and Elshaw cover long-standing issues. Another interesting feature of this volume concerns the extent of the international collaboration between authors. Two of the chapters are based on collaboration between authors from different countries.

Overall this volume reflects the diverse and dynamic nature of our field. We hope that readers will find something of interest in it.

CLC
ITR
May 2002

Chapter 1

FLEXIBLE WORKING ARRANGEMENTS: IMPLEMENTATION, OUTCOMES, AND MANAGEMENT

Suzan Lewis
Manchester Metropolitan University

Flexibility has become a buzzword in organizations. However, flexibility is an overarching term that incorporates a number of different types of strategy. Flexible working time and place arrangements, which are the subject of this chapter, are only one strand along with functional, contractual, numerical, financial, and geographical flexibility. This chapter focuses on flexible working arrangements (FWAs), that is organizational policies and practices that enable employees to vary, at least to some extent, when and/or where they work or to otherwise diverge from traditional working hours. They include, for example, flexitime, term time working, part-time or reduced hours, job sharing, career breaks, family-related and other leaves, compressed workweeks and teleworking. These working arrangements are also often referred to as family-friendly, work–family, or more recently work–life policies. This implies an employee focus, but the extent to which these policies primarily benefit employees or employers, especially in the 24/7 economy (Presser, 1998), or contribute to mutually beneficial solutions, has been the subject of much debate (e.g., Barnett & Hall, 2001; Hill, Hawkins, Ferris, & Weitzman, 2001; Purcell, Hogarth, & Simm, 1999; Raabe, 1996; Shreibl & Dex, 1998). Other work-family policies such as dependent care support can be used to complement FWAs and much of the research addresses them simultaneously. The term FWAs will be used in this chapter except where the research under consideration explicitly addresses work–family issues. Non-traditional work arrangements such as shift work or weekend work which are 'standard' in certain jobs are not considered here.

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

There are two increasingly converging strands of research on FWAs. One stems from a long tradition of examining flexible working as a productivity or efficiency measure (e.g., Brewster, Hegwisch, Lockhart, & Mayne, 1993; Dalton & Mesch, 1990; Krausz, Sagie, & Biderman, 2000) but increasingly also recognizes that these strategies have implications for work–personal life integration. The other has emerged from the work–life literature and depicts flexible working initiatives as tools for reducing work–family conflict or enhancing work–life integration, but has increasingly addressed productivity and other organizational outcomes (e.g., Barnett & Hall, 2001; Friedman & Greenhaus, 2000; Grover & Crooker, 1995; Hill et al., 2001; Kossek & Ozeki, 1998, 1999; Lewis, Smithson, Cooper, & Dyer, 2002; Prutchno, Litchfield, & Fried, 2000; Smith & Wedderburn, 1998). This review draws on literature from both traditions, although most studies are within the work–family paradigm. It focuses on three major current research themes: (i) empirical and theoretically based discussions of the factors contributing to organizational decisions to implement FWAs; (ii) research on the work-related outcomes of FWAs; and (iii) research focusing on issues in the management of flexible work and workers.

FACTORS ASSOCIATED WITH THE IMPLEMENTATION OF FWAs

Some forms of flexible working schedules such as part-time work, compressed work weeks, annualized hours, and flexitime have a long history and have traditionally been introduced largely to meet employer needs for flexibility or to keep costs down, though they may also have met employee needs and demands (Dalton & Mesch, 1990; Krausz et al., 2000; Purcell et al., 1999; Ralston, 1989). These and other flexible arrangements are also introduced ostensibly to meet employee needs for flexibility to integrate work and family demands under the banner of so-called family-friendly employment policies (Harker, 1996; Lewis & Cooper, 1995). Often a business case argument has been used to support the adoption of FWAs; that is, a focus on the cost benefits (Barnett & Hall, 2001; Bevan, Dench, Tamkin, & Cummings, 1999; Galinsky & Johnson, 1998; Hill et al., 2001; Lewis et al., 2002; Prutcho et al., 2000). Other contemporary drivers of change include increased emphasis on high-trust working practices and the thrust toward gender equity and greater opportunities for working at home because of new technology (Evans 2000). Nevertheless, despite much rhetoric about the importance of challenging outmoded forms of work and the gradual association of FWAs with leading-edge employment practice (DfEE, 2000; Friedman & Greenhaus, 2000; Friedman & Johnson, 1996; Lee, McDermid, & Buck, 2000), the implementation of these policies remains patchy across organizations (Glass & Estes, 1997; Golden, 2001; Hogarth, Hasluck, Pierre, Winterbotham, & Vivian, 2000). A major direction of recent research,

therefore, has been to examine the factors that influence organizational responsiveness to work–family issues and hence the development of FWAs.

This research initially emanated from North America (e.g., Goodstein, 1994; Ingram & Simons, 1995; Milliken, Martins, & Morgan, 1998; Osterman, 1995) but also includes some recent research from Europe and Australia (Bardoel, Tharenou, & Moss, 1998; den Dulk, 2001; Dex & Schreibl, 2001; Wood, De Menezes, & Lasaosa, forthcoming). It has focused on identifying factors associated with the adoption of formal FWAs and other work–family policies rather than actual practice and employee use of these initiatives. Organizational size, and sector and economic factors are widely identified as being associated with the adoption of policies (Bardoel et al., 1998; den Dulk & Lewis, 2000; Goodstein, 1994; Ingram & Simons, 1995; Milliken et al., 1998; Wood, 1999; Wood et al., forthcoming). The research suggests that large organizations are more likely to provide formal FWAs than smaller ones; public sector organizations are more likely to develop initiatives than private sector companies; and, within the private sector, arrangements are more common in the service and financial sector compared with construction and manufacturing (Bardoel et al., 1998; Forth et al., 1997; Hogarth et al., 2000; Ingram & Simons, 1995; Morgan & Milliken, 1992). These sectors employ more women, and it is usually believed that having more women in the workforce creates internal pressures that are associated with the development of work–family polices. However, findings on the influence of the proportion of women in the workforce are mixed. Some studies find this factor is associated with the likelihood of adopting FWAs and work–family policies such as childcare (Auerbach, 1990; Bardoel et al., 1998; Glass & Fujimoto, 1995; Goodstein, 1994), while this relationship is not found in other studies (Ingram & Simons, 1995; Morgan & Milliken, 1992). This may depend on the position of women as there is evidence that organizations with a relatively large share of women managers seem to provide work–family arrangements more often than organizations where women's employment consists mainly of lower skilled jobs (Glass & Fujimoto, 1995; Ingram & Simons, 1995). However, when access to flexible work schedules rather than work–family policies (which include dependent care and family related leaves) are considered, women are less likely than men to have access to them (Golden, 2001). Other research suggests that organizations with relatively 'progressive' employment policies and philosophies, seeking to implement high-commitment management, may also be likely to develop FWAs and other work–family supports (Auerbach, 1990; Osterman, 1995; Wood et al., forthcoming).

Theoretical Frameworks

The majority of studies in this tradition have been based on the analysis of large-scale surveys of policies implemented in organizations, usually testing

predictions derived from variations on institutional theory (Goodstein, 1994; Ingram & Simons, 1995; Kossek, Dass, & DeMarr, 1994; Ingram and Simons, 1995; Morgan et al., 1998). The institutional theory approach begins with the basic assumption that there is growing institutional pressure on employers to develop work–family arrangements. It is argued that changes in the demographics of the workforce have increased the salience of work–family issues, and that public attention to these issues and/or state regulations have heightened institutional pressures on employers to be responsive to the increasing need for employees to integrate work and family demands. Variability in organizational responses to these normative pressures is explained by differences in the visibility of companies and in the extent to which social legitimacy matters to them. Hence public sector and large private sector organizations are most likely to develop policies because of concern about their public image. Pressure is also exerted when other organizations in the same sector introduce flexible policies (Goodstein, 1994). Critics of the traditional, institutional theory approach maintain that this underestimates the latitude available to employers to make strategic decisions in adapting to institutional pressures. Goodstein (1994) argues that responsiveness to institutional expectations depends on both the strength of institutional pressures and on economic or other strategic business or technical factors, such as the need to retain skilled staff and the perceived costs and benefits of introducing work-family arrangements. More recently a number of variations of institutional theory and other theoretical approaches have been proposed, differing in the extent to which they focus on institutional pressures, organizational factors, and technical or business considerations (Wood et al., forthcoming). Recent attempts to identify significant factors associated with the adoption of policies, however, suggest that, while all theoretical approaches have some value, no single theoretical perspective can explain all the findings (Wood et al., forthcoming; Dex & Shreibl, 2001). Institutional pressures, strategic business concerns, local situational variables, and human resource strategies may all influence organizational decision making to some extent.

Two major limitations of research examining the factors associated with organizational responsiveness to work–life issues (and indeed much of the other literature in this area) have been the tendency to focus on large organizations and on formal policy rather than informal practice. There is a growing consensus that the availability of formal FWAs alone is not necessarily indicative of their use in practice (e.g., Cooper, Lewis, Smithson, & Dyer, 2001; Lee et al., 2000; Lewis et al., 2002; Rapoport, Bailyn, Fletcher, & Pruitt, 2002), and this is discussed in later sections of this chapter. The neglect of small- and medium-sized organizations also relates to this policy/practice distinction. The scope for informal practices and flexibility in smaller organisations is often overlooked.

Small- and medium-sized organizations

There is some indication that smaller organizations are more likely than larger ones to develop informal practices, which are often implemented in an *ad hoc* way, to meet the needs of individual employees (Bond, Hyman, Summers, and Wise, 2002; Cooper et al. 2001; Dex & Schreibl, 2001), although one survey failed to confirm this (MacDermid, Litchfield, and Pitt-Catsouphes, 1999). Findings that the size of companies is a predictor of FWAs may thus be an artefact of what it is that is measured. More informal FWAs may well be more appropriate for small- and medium-sized organizations because of their fewer resources and their greater difficulty in, for example, getting cover for colleagues on leave or working flexible hours. However, as with larger companies, no single theoretical approach appears to explain why FWAs are implemented in small- and medium-sized enterprises. For example, Dex and Shreibl (2001) describe a range of formal and informal arrangements that were introduced in small businesses in response to institutional, business, and economic pressures as well as ethical concerns. They found that small organizations were more hesitant about introducing flexibility and were particularly concerned about costs; but it was also in small businesses compared with large businesses in their study that attempts were made to introduce a culture of flexibility (e.g., encouraging employees to cover for each other). Lack of formalization of policies in small businesses could be associated with inequity. On the other hand, formal policies in larger organizations are not necessarily applied in an equitable or consistent way (Bond et al., 2002; Cooper et al., 2001; Lewis, 1997; Lewis et al., 2002; Powell & Mainiero, 1999), and there is some evidence that employees in small organizations with informal practices can feel more supported than those in large organizations with a coherent programme of policies but difficulties in practice (Cooper et al., 2001).

The role of national social policy and state legislation

Recent research, particularly European and cross-national studies, have begun to examine the processes whereby social policy and state legislation might influence the adoption of workplace policies (den Dulk, 2001; Evans, 2000; Lewis et al., 1998). Social policy, such as the statutory provision of childcare and legislation to support work and family integration, varies cross-nationally. For example, paid parental leave is an entitlement in many European states, and in some countries, particularly in Scandinavia, fathers as well as mothers are encouraged to take up this entitlement (Brannen, Lewis, Nielson, & Smithson, 2002; Moss & Deven, 1999); maternity but not parental leave (for either parent) is paid in the UK and parental leave is unpaid in the USA. Employees are more likely to take up parental leave entitlements if they are remunerated (Moss & Deven, 1999), so

organizations develop voluntary FWAs, especially in relation to leave, in different contexts. In Europe state legislation requires organizations to implement policies such as parental leave, the right to take leave for family emergencies, or the provision of equal pro rata benefits for part-time workers. Legislation may also help to create a normative climate that gives rise to higher expectations of employer support (Lewis & Lewis, 1997; Lewis & Smithson, 2001). Edelman argued, 'when a new law provides the public with new expectations or new bases for criticising organisations, or when the law enjoys considerable societal support, apparent non-compliance is likely to engender loss of public approval' (Edelman, 1990, p. 1406). Social policies such as the provision or absence of publicly provided childcare also contribute to institutional pressures on organizations to take account of work-family issues. Evidence from a five-country European study of young workers' orientations to work and family suggests that supportive state policies including legislation and public childcare provision can enhance young people's sense of being entitled to expect support for managing work and family, not just from the state but also from employers (Lewis & Smithson, 2001), which may increase internal as well as external pressures on organizations.

There has been some debate about whether statutory entitlements and provisions encourage employers to implement more voluntary FWAs and other work-family policies, which would be in keeping with institutional theory, or whether it absolves them from responsibility for employees' non-work lives, which might suggest that economic factors are more important (Brewster et al., 1993; Evans, 2000). An overview of analyses of provisions in EU countries suggests that voluntary provision by companies are highest in countries with a medium level of statutory provision such as Austria and Germany. They are least likely to be implemented in those countries with the lowest levels of statutory provision such as the UK and Ireland and in those with the highest levels of support also; that is, the Nordic countries (Evans, 2000). One explanation of this finding may be that national legislation tends to encourage private provision up to a point, after which it tends to replace it, although Evans cautions that it is also necessary to take account of the possible impact of cultural attitudes toward the family on both public policy and the behavior of firms. Another possible explanation for the finding that high levels of statutory provision appear to be associated with lower employer provision may be that national surveys of employer policies tend to focus on childcare support, and on family leaves beyond the statutory minimum, to a greater extent than flexible forms of work. Dependent care policies are less relevant in, for example, the Nordic countries where public provision of childcare is high and statutory leave rights are generous. Elsewhere employers may introduce voluntary provisions to compensate for lack of state provision (den Dulk & Lewis, 2000), while employers in countries with a higher level of statutory provision

do not have to give so much consideration to providing support for childcare or parental leaves and are therefore free to focus on flexible ways of organizing work. Indeed, it has been suggested that the need to organize work to accommodate family leaves can oblige employers to develop such flexibility (Kivimaki, 1998). More cross-national studies focusing specifically on FWAs will be necessary to examine this possibility. Cross-national studies focusing on the development of good practice from the employees perspective rather than policies as reported by organizations would elucidate further the impact of national policy.

Organizations thus implement flexible and work–family arrangements in response to internal and external pressures although technical factors are also taken into account. One of the most influential technical factors is the business case; that is, the argument that the development of FWAs is cost-effective and has a positive impact on recruitment, retention, turnover, and other work-related variables (Bardoel et al., 1998; Bevan et al., 1999; Prutchno et al., 2000). But how viable is this argument? Much of the human resource (HR) literature that sets out the business case is based on small-scale or large-scale but organizationally specific case studies (e.g., Bevan et al., 1999; DfEE, 2000; Hill et al., 2001). It also fails to consider the possibility that if FWAs have no costs (rather than an actual bottom-line benefit) there may still be an important case for their implementation (Harker & Lewis, 2001). Another theme of recent research in this area has been to examine more critically the organizational outcomes of flexible working practices.

THE OUTCOMES OF FLEXIBLE WORKING ARRANGEMENTS

Evaluation studies vary in the FWAs that they address, the methods they use, and the outcome variables studied. Furthermore, even when the same outcome variables are employed different measures are often used, making comparisons difficult. Nevertheless research on the outcomes of FWAs demonstrates that, although there can be some positive work-related outcomes, a simple business case argument neglects much of the complexity in this area.

Outcomes vary by types of FWA and outcome studied

Numerous studies and several recent reviews and meta-analyses of outcome research have concluded that flexible working arrangements can have positive organizational effects, at least in some circumstances (Friedman & Greenhaus, 2000; Baltes, Briggs, Huff, Wright, & Neuman, 1999; Kossek & Ozeki, 1999; Glass & Estes, 1997; Hill et al., 2001), although the results are

not always consistent and reported outcomes are sometimes minimal and often contingent upon other factors. Kossek and Ozeki (1999) carried out a meta-analysis of studies examining (a) the relationships between work-family conflict and organizational outcomes, (b) work–family policy (including both FWAs and dependent care policies) and organizational outcomes, or (c) the overall links between policies, conflict, and outcomes. Criteria for inclusion of studies were that they reported a correlation between a work–family conflict measure and one of six work-related outcomes (performance, turnover intentions, absenteeism, organizational commitment, job/work involvement, and burnout) or that they estimated the effects of an HR policy or intervention on one of the six work-related outcomes or work–family conflict. They found qualified support for the relationship between policies and the work-related outcomes although the results varied somewhat according to the policies and outcomes studied. FWAs tended to be more strongly related than dependent care measures to work-related outcomes but policies did not necessarily reduce work-family conflict nor improve organizational effectiveness in all circumstances, particularly if they did not enhance employees' sense of control over their work schedules.

Many of the studies reviewed by Kossek and Ozeki (1998) were cross-sectional in design so that effects over time were not clear. Baltes et al. (1999) carried out a meta-analysis of the effects of experimental intervention studies of flexitime and compressed work weeks selecting only those studies that included pre- and post-intervention test measures or normative experimental comparison and found that results varied according to the policy and outcomes assessed as well. The meta-analysis was theory driven with hypotheses derived from a range of theoretical models including the work adjustment model, job characteristics theory, person–job fit, and stress models. Baltes et al. (1999) concluded that both flexitime and compressed work weeks had positive effects on productivity/or self-rated performance, job satisfaction, and satisfaction with work schedules but that absenteeism was affected by flexitime only. They suggest that the different effects on absenteeism are because compressed work weeks are less flexible and therefore do not allow employees to, for example, make up time lost through illness or other reasons, as flexitime does. However, this appears to contradict another of their findings, namely that the degree of flexibility is negatively associated with the organizational outcomes studied (Baltes et al., 1999). This suggests that too much flexibility is a bad thing. This finding, which is both counter-intuitive and also counter to theory-based predictions, is explained by Baltes et al. in terms of the possible difficulties in co-ordinating and communicating with others that might arise if there is too much flexibility. However, it is worth noting that, although their analysis is relatively recent, the studies examined in this meta-analysis were mainly conducted in the 1970s and 1980s, many of them before the recent developments in information and communication technologies that are so crucial for many forms of flexible

working. In the age of mobile phones and emails, communication difficulties may be much less pertinent. Other more recent research, albeit not experimentally based, suggests the opposite: that more rather than less flexibility is associated with more positive outcomes, at least in terms of self-reported outcomes. For example, Prutchno et al. (2000), in a survey of over 1,500 employees and managers in six US corporations, found that daily flexitime, which they defined as schedules that enable employees to vary their work hours on a daily basis, was much more likely than traditional flexibility to be associated with self-reported positive impacts on productivity, quality of work, plans to stay with the company, job satisfaction, and a better experience of work–family balance. Other studies have suggested that the impact of flexible working arrangements on organizational outcomes may depend less on the objective extent of flexibility than on psychological factors such as preferred working schedules (Ball, 1997; Barnett, Gareis, & Brennan, 1999; Krausz et al., 2000; Martens, Nijhuis, van Boxtel, and Knottnerus, 1999) or the extent to which flexibility provides autonomy and control (Tausig & Fenwick, 2001; Thomas & Ganster, 1995), as discussed later in this chapter.

The possibility of having too much flexibility is, however, raised in research focusing on teleworking, that is, working from home for some or all the week, which has produced mixed results. There is some evidence of positive work-related outcomes such as higher job satisfaction, organizational commitment, and lower turnover among teleworkers than office-based workers and of enhanced flexibility and integration of work and non-work roles in some circumstances (Ahrentzen, 1990; Dubrin, 1991; Frolick, Wilkes, & Urwiler, 1993; Olsen, 1987; Rowe & Bentley, 1992). However, other research reports negative outcomes such as lower job satisfaction and organizational commitment, less positive relationships with managers and colleagues, greater work–family conflict and more tendency to overwork such as working during vacations (Olsen, 1987; Prutchno et al., 2000). Many studies imply that teleworking can be a double-edged sword with the potential for both positive and negative outcomes (Hill, Hawkins, & Miller, 1996; Steward, 2000; Sullivan & Lewis, 2001). The tendency to overworking may be regarded as symptomatic of the increased blurring of work and non-work boundaries that appears to be becoming widespread, facilitated by developments in information and communication technologies (Haddon, 1992; Steward, 2000; Sullivan & Lewis, 2001). Although the outcome measures used in research on telework often differ from those used in relation to other forms of flexible working arrangements, research does seem to suggest that too much flexibility in the context of information and communication technology in the home as well as the workplace may raise a different set of issues about impacts on individuals, their families, and organizations that will require further exploration and research (Standen, Daniels, & Lamond, 1999). The effects of teleworking also appear to be highly gendered. Qualitative research shows that teleworking women are

more likely than men to multitask and less likely to have a room of their own, men more likely to be able to shut themselves away and work without interruptions from family members (Sullivan, 2000; Sullivan & Lewis, 2001). This may explain why, in a recent UK national survey, men were more likely than women to wish to work from home (Hogarth et al., 2000).

The different dependent variables used in outcome research makes comparison difficult. This is particularly evident in relation to measures of performance and productivity. Measures used in the research include sales performance (Netemeyer, Boles, & McMurrian, 1996), self-rated performance (Cooper et al., 2001; Prutchno et al., 2000), self-efficacy ratings (Kossek & Nichol, 1992; Netemeyer et al. 1996), and supervisor ratings (Kossek & Nichol, 1992), as well as more objective measures of productivity in the case of manufacturing workers (Baltes et al., 1999; Shepard & Clifton, 2000). It can be argued that studies that predetermine outcomes inevitably limit to some extent what can be learnt about FWAs. Action research, which begins at an earlier stage with the problem to be solved rather than the evaluation of policy to be implemented, may have advantages in this respect. Rather than just predetermining what outcomes will be measured, this approach enables other outcomes to emerge, grounded in the specific organizational context. For example, action research carried out in a number of companies in the USA explored systemic solutions that could meet both strategic business needs and employees needs for work–life integration as well as gender equity—what the researchers term the dual agenda (Bailyn, Rapoport, Kolb, and Fletcher, 1996; Fletcher & Rapoport, 1996; Rapoport et al., 2002). Interventions developed as a consequence of the research team working collaboratively with employees, examining the nature of the work, and identifying, surfacing, and challenging assumptions about working practices, included introducing periods of quiet time so that work could be carried out without interruptions, and removing management discretion about FWAs in order to empower work teams to develop their own schedules. It is worth noting that the outcomes identified from these interventions were both more varied and more directly bottom-line-oriented than those in most experimental or survey research. They were specific to the work unit studied and included improved time to market, enhanced product quality, and increased customer responsiveness as well as more traditional measures such as reduced absenteeism (Fletcher & Rapoport, 1996; Rapoport et al., 2002).

Processes and Intervening Variables

The action research discussed above focused on evaluation of the process of bringing about change rather than policy. In contrast, much of the research evaluating FWAs neglects process, although this is crucial for explaining how and why some FWAs are effective in some circumstances. There have, never-

theless, been a number of attempts to theorize the outcomes of FWAs or work-life policies and the factors which facilitate or undermine them. Research has examined the role of work–family conflict, worker preferences, perceived control and autonomy, perceptions of organizational justice, perceived management and organizational support, and organizational learning. These studies tend to address the processes explaining outcomes of FWAs for those with family commitments, primarily childcare and eldercare demands. Less attention has been paid to the processes whereby FWAs may impact on work-related outcomes among employees more broadly or on more fundamental organizational change.

Work–Family Conflict

It is often argued that FWAs can contribute toward positive integration of work and personal life (Galinsky & Johnson, 1998; Hill et al., 1996). However, much of the research operationalizes this in terms of absence or minimization of work–family conflict. Kossek and Ozeki (1999) argue that work–family conflict is a crucial but often neglected variable for understanding the process whereby FWAs may relate to work-related outcomes. Studies examining work–family conflict increasingly distinguish between work conflicting with family and family conflicting with work rather than more global measures (e.g., Burke & Greenglass, 2001; Frone, Russell, & Cooper, 1992; Frone, Yardley, & Markel, 1997; Kelloway, Gottlieb, & Barham, 1999; Kossek & Ozeki, 1998, 1999; Netemeyer et al., 1996). Kossek and Ozeki (1999) conclude from their meta-analysis that, although work conflicting with the family role is not necessarily related to productivity and work-related attitudes, there is substantial evidence that family conflicting with the work role is. To understand why and how FWAs influence individual work-related attitudes and behaviours, they argue, it is necessary to examine how they affect different aspects of work–family conflict, which in turn influence outcomes such as performance and absenteeism. For example, there may be different implications of the effects of FWAs on time-related strains or emotional conflict: 'not being able to do two things at the same time may impact differently from feeling bad about it!' (Kossek & Ozeki, 1999, p. 18). As Kossek and Ozeki (1999) point out there is a need for more longitudinal research looking at the impact of FWAs on work–family conflict to extend our understanding of the processes whereby policies impact on individual and organizational outcomes. The impacts of FWAs on work–family conflict and subsequent work-related outcomes may vary for different groups of workers and this too needs to be taken into account in this research. Gender is a crucial variable (Greenhaus & Parasuraman, 1999) as well as the gender composition of workplaces (Holt & Thaulow, 1996; Maume & Houston, 2001). Age (or generation) may be a further relevant factor. For example, in a study of chartered accountants in the UK, the link between

work–family conflict and intention to leave was particularly strong among the younger generation (Cooper et al., 2001), who were also more likely to say they would use FWAs.

Employee Preferences

As would be predicted from person–job fit theory, the impact of FWAs appears to depend on employee preferences (Ball, 1997; Krausz et al., 2000; Martens et al., 1999). The work-related outcomes can be positive if FWAs fit in with employee needs but may be non-effective or even detrimental if not freely chosen (Martens et al., 1999; Tausig & Fenwick, 2001). Martens et al. (1999) concluded that FWAs were only beneficial to employees who could choose and control their own flexibility, after finding significantly more health problems among Belgian employees working flexible schedules than among a control group working traditional hours. However, the FWAs examined in this study included long or irregular shifts, on-call work, and temporary employment contracts, implemented for employer flexibility.

Control and Autonomy

A closely related factor influencing the outcomes of FWAs is the extent to which they are perceived as providing control and autonomy over working hours (Krausz et al., 2000; Martens et al., 1999; Tausig & Fenwick, 2001; Thomas & Ganster, 1995). Thomas and Ganster (1995) distinguished between family supportive policies and family supportive managers, both of which they found related to perceived control over work and family demands, which, in turn, were associated with lower scores on a number of indicators of stress among a sample of healthcare professionals. While there is much support for the view that perceived control and autonomy explain positive outcomes of some FWAs (Dalton & Mesch, 1990; Kossek & Oseki, 1999; Tausig & Fenwick, 2001), this does seem to depend to some extent on the populations studied. Baltes et al. (1999) found that the benefits of flexitime and compressed work weeks were lower for managers than other employees and argue that this is likely to be because managers already have considerable autonomy and therefore these polices are less relevant. Other research however, contests the idea that management autonomy provides flexibility. It is often more difficult for managers to work flexibly, particularly in the context of long-working-hours cultures (Kossek et al., 1999; Perlow, 1998; Bond et al., 2002; Hochschild, 1997; Lewis et al., 2002) and the beliefs that managerial or supervisory tasks cannot be performed flexibly (Powell & Mainiero, 1999). In fact a theme in much current research is that those workers who have opportunities to work flexibly and have autonomy to manage their own work schedules often use this to work longer rather than shorter hours (Holt & Thaulow, 1996; Lewis et al., 2002; Perlow, 1998).

More research is needed to clarify the effects of FWAs on managers and subsequent organizational outcomes if the effects are, perversely, to encourage or enable managers and professionals to work longer hours (Lewis & Cooper, 1999).

Perceived Organizational Justice

Although FWAs can potentially benefit all employees and their employing organizations, they are often directed mainly at employees with family commitments, especially parents of young children (Young, 1999). There have been some suggestions, mostly in the popular media, of work–family backlash among employees without children, especially if they feel that they have to do extra work to cover for colleagues working more flexibly (Young, 1999; Lewis et al., 1998). This raises the possibility of negative organizational outcomes of FWAs in some circumstances. These might include low morale or resentment, which, in turn, could affect job satisfaction, intention to leave, and other outcomes among employees who do not have access to FWAs, if this is selectively provided. Results from the sparse research that has addressed these questions are inconsistent. Grover (1991) examined perceptions of fairness of family-related leave in the USA and concluded that these were influenced by whether or not employees were likely to gain personally. Thus parents and those who were considering becoming parents viewed these leaves more favourably than other employees, supporting the backlash notion. In contrast, Grover and Crooker (1995) compared employees in organisations with and without family-oriented policies and found that all employees in family responsive organizations, regardless of their own parental status, perceived their employers more positively. The authors suggested that work–family policies contributed to perceptions of the organization as being generally supportive and fair, which contradicts the idea of backlash. Parker and Allen (2000) extended this research by examining a number of personal and situational factors that might impact on perceptions of fairness of work–family policies and generated moderate support for both views. Employees who had personal experience of using FWAs, younger employees, women, and parents with young children (but not those who were considering becoming parents) were most likely to perceive work–family policies as fair. The only situational variable to influence fairness perceptions was interdependence of tasks. It was expected that employees working in jobs characterized by high job interdependence would be more likely to perceive work–family policies as unfair because they would be more inconvenienced by colleagues flexibility. The findings were significant in the opposite direction. The authors speculate that this may be due to the personality characteristics of those who self-select into this type of job, and who may be more relationship oriented. Another possible explanation is that employees in highly interdependent jobs may be more aware of the

potential reciprocity of informal and formal flexibility from which they too could benefit (Holt & Thaulow, 1996).

In circumstances where FWAs are directed primarily at parents, perceived inequity among employees without children can also reduce the sense of entitlement to take up provisions among parents themselves (Lewis, 1997; Lewis & Smithson, 2001), reducing the takeup and therefore outcomes of FWAs. Making FWAs normative and available to all rather than subject to management discretion may therefore contribute more than targeted policies to a family supportive culture.

The impact of FWAs can also be influenced by perceived procedural justice. Interventions in which employees have been able to participate in the design of work schedules appear to have the potential to achieve highly workable, flexible arrangements and be associated with positive work-related attitudes (Ball, 1997; Kogi & Martino, 1995; Rapoport et al., 2002; Smith & Wedderburn, 1998). Conversely, lack of consultation with managers about the development of FWAs can contribute to feelings of unfairness, which may undermine implementation. For example, Baltes et al. (1999) speculated that one reason for the negative effects of high levels of flexibility in the studies they reviewed may be that more flexible policies on paper may result in managers clamping down on flexibility in practice in order to sustain control. Dex and Schriebl (2001) also noted that some of the managers in the larger organizations they studied felt alienated because they were compelled to introduce policies on which they had not been consulted.

Clearly more research is needed to clarify conditions under which FWAs are perceived as fair, the outcomes of justice perceptions, and the implications for organizations. Organizational justice theory (e.g., Greenberg, 1996) provides a useful framework for extending understanding of these processes (Lewis & Smithson, 2002; Young, 1999).

Organizational Culture and Supportive Management

Organizational culture or climate is a crucial variable contributing to the outcomes of FWAs, especially when these are formulated as 'family friendly' rather than productivity measures (Bailyn, 1993; Fried, 1998; Friedman & Johnson, 1996; Hochschild, 1997; Lewis, 1997, 2001; Lewis et al., 2002). In this context aspects of culture such as the assumption that long hours of face time in the workplace are necessary to demonstrate commitment and productivity, especially among professional and managerial workers, can co-exist with more surface manifestations of work–life support (Bailyn, 1993; Cooper et al., 2001; Lewis, 1997, 2001; Lewis et al., 2002; Perlow, 1998; Rapoport et al., 2002). Moreover, opportunities for flexible working are not always well communicated (Bond et al., 2002). Often employees with most need for flexibility are unaware of the possibilities (Lewis, Kagan, & Heaton, 1999). Supervisory support is a critical aspect of the organizational climate that is

essential for policies to be effective in practice (Goff, Mount, & Jamison, 1990; Thomas & Ganster, 1995), but it is not always forthcoming and many employees feel that taking up opportunities for flexible working will be career limiting (Bailyn, 1993; Cooper et al., 2001; Lewis et al., 2002; Perlow, 1998). In many occupations, especially at professional and managerial levels 'strong players' are regarded as those who do not need to modify hours of work for personal reasons (Lewis, 2002).

The impact of workplace culture on the outcomes of FWAs has not always been acknowledged in discussions of notions such as family friendliness. Initially, family friendliness was measured by the number of policies available (Galinsky, Friedman, & Hernandez, 1991). Wood (1999) proposed a concept of family-friendly management that denotes a coherent rather than *ad hoc* approach to the development of work–family policies. This implies more consistency in supportiveness for employees with family commitments, but still tends to be measured by reference to policy adoption as reported by HR representatives or other managers (Wood, 1999). To understand the process whereby FWAs may relate to work-related behaviors and attitudes it is important to understand the organizational climate from employees perspectives. Recent literature has begun to focus more on organizational culture, and measures to assess the extent to which organizational cultures are perceived as supportive of work–family integration have been developed (Allen, 2001; Lyness, Thompson, Francesco, & Jusiesch, 1999; Thompson, Beauvais, & Lyness, 1999). Drawing on theories of organizational and social support Thompson and her colleagues (Thompson et al., 1999) developed a measure of perceived organizational family support (POFS) that assesses perceived instrumental, informational, and emotional support for work–family needs. It incorporates perceived support from the organization and from supervisors. Another scale (Allen, 2001) examines global employee perceptions of the extent to which their organizations are supportive. Both measures predict work-related outcomes in the expected direction, including enhanced organizational commitment, job satisfaction, women's intentions to return to work more quickly after childbirth and reduced intention to quit, and work–family conflict (Allen, 2001; Lyness et al., 1999; Thompson et al., 1999), and have been found to mediate the relationship between FWAs and work-related behaviours and attitudes. The distinction between perceived support from the organization and from supervisors may be a fruitful avenue to pursue further. In some cases wellintentioned support from the organization (i.e., senior or HR management) can be undermined by line managers (Lewis & Taylor, 1996) while there is some evidence that line managers can be perceived as more supportive than 'the organization' when the normative culture is not perceived as supporting work–life integration (Cooper et al., 2001).

The inclusion of measurements of employees perceptions of organizational climate in studies evaluating FWAs is an important advance, recognizing the

distinction between policy and practice and underscoring the fact that FWAs will have limited impact if not supported by the culture. However, future research could usefully generate multiple perspectives, including managers' perceptions of FWAs in practice, to provide a more holistic picture of the processes and barriers impacting the effectiveness of FWAs in a range of workplace contexts. Non-supportive cultures suggest that understanding of the potential value of FWAs has not been diffused throughout the organization and this has not become a part of organizational learning.

FWAs and Organizational Level Change: Organizational Learning

Although it is increasingly recognized that FWAs can meet the needs of both organizations and individual employees, most research on FWAs has not only focused on policy rather than practice but also on individual level outcomes rather than organizational level change. Some recent research, however, has begun to focus on the organizational level of analysis, examining the contribution of FWAs to organizational learning and change (Lee et al., 2000; Rapoport et al., 2002). Lee et al. (2000) examined responses to managerial and professional workers' requests for reduced-hours work in terms of the organizational learning that takes place. They found three different paradigms of organizational learning in this situation: accommodation, elaboration, and transformation. Accommodation involves making individual adaptations to meet the needs of specific employees, usually as a retention measure, but not involving any broader changes. Indeed, efforts are made to contain and limit this different way of working, rather than using this as an opportunity for developing policies or broader changes in working practices. In other organizations with formal policies on FWAs, backed up by a well-articulated view of the advantages to the organization, *elaboration* takes place. This goes beyond random individual responses to request for flexibility, but full-time employees are still the most valued as employers make efforts to contain and systematize procedures for experimenting with FWAs. In the *transformation* paradigm of organizational learning FWAs are viewed as an opportunity to learn how to adapt managerial and professional jobs to the changing conditions of the global marketplace. The concern of employers is to experiment and learn. These emergent paradigms were considered by Lee et al. (2000) to be representative of more general organizational variability in response to changes in the external environment or challenges to the status quo.

The notion that FWAs can be a strategy for responding to key business issues implicit in the transformation paradigm is also highlighted in studies that have employed action research to bring about organizational change to meet a dual agenda of organizational effectiveness, on the one hand, and work–personal life integration and gender equity, on the other (Fletcher & Rapoport, 1996; Rapoport et al., 2002). This approach illustrates a process

whereby organizational learning can be deliberately helped along. Managers, at all levels, are crucial in this process. The next section, ‘‘The management of flexible work and workers’’, therefore examines research on the role of managers in the implementation and diffusion of FWAs.

THE MANAGEMENT OF FLEXIBLE WORK AND WORKERS

An important but relatively new area of research concerns how managers make decisions about the day-to-day operation of FWAs. It is clear from both qualitative and quantitative research that management attitudes, values, and decisions are crucial to the effectiveness of FWAs (Bond et al., 2002; Dex & Schreibl, 2001; Goff et al., 1990; Hochschild, 1997; Lee et al., 2001; Lewis, 1997, 2001; Perlow, 1998; Rapoport et al., 2002; Thompson et al., 1999). Managers must communicate, implement, and manage FWAs within organizational cultures that they both influence and are influenced by. Managers can increase the effectiveness of FWAs by their supportiveness (Allen, 2001; Hohl, 1996; Thomas & Ganster, 1995; Thompson et al., 1999) or can undermine FWAs by communicating, in a variety of ways, implicit assumptions about the value of more traditional ways of working (Lewis, 1997, 2001; Perlow, 1998). Managers also influence flexible working by their response to requests for non-standard work, the ways in which they manage flexible workers on a day-to-day basis, and by their own flexibility and work–life integration.

Management Decision Making in Relation to Subordinates Requests for Flexible Work Schedules

FWAs are often subject to management discretion; that is, first line managers with operational responsibility have the discretion to say who can work in flexible ways. Managers decisions to grant requests may be based on beliefs about potential disruption, substitutability of employees, notions of fairness and respect, perceptions of employees’ record of work and commitment, perceived long-term impact, or perceived gender appropriateness of a request and other factors (Bond et al., 2002; Klein, Berman, & Dickson, 2000; Lee et al., 2000; Powell & Mainiero, 1999). Powell and Mainiero (1999) analysed managers’ decision making when responding to vignettes in which hypothetical subordinates made requests for FWAs (working from home for part of the week, part-time work, or unpaid leave). The type of FWA requested, characteristics and job role of the subordinate, and reasons for the request were all manipulated. They found support for a work disruption theoretical explanation of first line managers’ decision making. That is, managers looked more favorably on requests that they perceived as involving least disruption. For example, working from home

was regarded more favourably than unpaid leave, and managers tended to be less willing to approve FWAs for subordinates whose skills and tasks were critical to operations and could not be easily replaced or who had supervisory responsibilities. This reflects other research that suggests that managers may be more willing to grant requests for flexibility to those workers who are more easily substituted (Bond et al., 2002). The tendency for decisions to be made on the basis of judgements of short-term disruption rather than long-term consideration of the potential costs of losing and replacing subordinates who are most critical to the work unit suggests that many of the managers are not aware of, or do not fully understand, long-term arguments for flexible working. However, Powell and Mainiero (1999) found some diversity in decision making, particularly in relation to the extent to which managers focused on the person making the request, the nature of the FWA requested, and the reason for making the request. In so far as this mirrors actual behavior in organizations this implies that decision making is not always based on consistent principles, which can have important implications for employees' perceptions of justice and possible backlash.

In contrast, however, other policy capturing research suggests that managers may be more likely to grant requests for alternative work schedules to those on whom they rely most. Klein et al. (2000) asked a sample of American attorneys (including both partners and associates) to rate how likely it was that their firm would grant requests from different attorneys to change from full-time to part-time work, again using hypothetical scenarios. They drew on dependency theory (Bartol & Martin, 1989) to predict that managers would be most likely to acquiesce to those subordinates on whom they most depended and on institutional theory to predict that managers would respond more favourably to requests from women and when requests related to childcare than for other reasons. Both hypotheses were supported. The notion of dependency is conceptually and operationally similar to that of criticality of subordinates used by Powell and Mainiero (1999) but also differs in notable ways. Both incorporate the ease or difficulty of replacement but the scenarios used in Klein et al.'s (2000) study also included high performance, the subordinates having support from powerful people in the organization and subordinates' threats to leave. It may be that if these extra variables had been included in Powell and Mainiero's study managers would have been more reflective in responding to requests. However, there are other differences between the two studies, which may also explain the contradictory findings about critical employees. Klein et al. studied lawyers while Powell and Mainiero looked at managers in a range of organizations. Powell and Mainiero also included a wider range of alternative working strategies. Perhaps most significant is the distinction between manager supportiveness and perceived organizational supportiveness (Allen, 2001). Powell and Mainiero examined managers' decision making processes while Klein et al. addressed the perspectives of both managers and subordinates on

their firms' likely responses. Managers, especially HR managers, tend to present different and often more favourable pictures of alternative working strategies in the organizations than subordinates (Klein et al., 2000; Cooper et al., 2001). Thus the contradictory findings may indicate contradictions between what managers say they would do and more general perceptions of organizational responsiveness.

The two studies also produced contradictory findings on the significance of the gender of subordinates making the request. Klein et al.'s (2000) findings support institutional theory and other research on the role of applicants gender (Barham, Gottlieb, & Kelloway, 1998) in that managers were more likely to grant requests to women than men and to those related to childcare than those from employees who wanted time for writing a novel. Powell and Mainiero (1999), on the other hand, found managers less willing to grant requests relating to childcare than eldercare, which was viewed as short term and therefore potentially less disruptive. In their study, the gender of the subordinate did not significantly affect decisions though the gender of the manager did, with women being more likely to make favourable decisions. While the difference between these two sets of findings may again be related to methodological approaches, it is possible that managers may be less influenced by gender stereotypes than others believe. If so this could have a significant effect. If men, and also women who want flexibility for non-childcare-related reasons, believe that managers are unlikely to grant requests for FWAs this may reduce the likelihood of making such requests and therefore reduce the scope and pressure for managers to make non-stereotyped decisions.

Both studies are limited in that they use hypothetical scenarios and methods that restrict the number of variables that can be manipulated and analysed. Given the contradictory nature of these findings more research is needed to clarify the factors that impinge on management decision making about FWAs, including the influence of the likelihood of different groups of subordinates actually making such requests, in a range of real-life organizational settings and the implications for organizational learning.

Management Expectations of Flexible Workers

A second way in which managers can influence the effectiveness of FWAs is by their day-to-day management and expectations of flexible workers and those who have access to FWAs. There is some evidence from both qualitative and quantitative research suggesting that employees working shorter hours may be as or often more efficient or productive than full-timers (Lewis, 1997, 2001; Stanworth, 1999). Yet studies of part-time and reduced-hours workers suggest they often pose particular problems for managers, particularly in contexts where those working non-standard hours are in a minority, the result of reactive decision making rather than part of a

well-thought-out strategy, and in the context of a norm of long working hours (Cooper et al., 2001; Edwards & Robinson, 2000; Lewis, 2001; Lewis et al., 2002). Managers do not always adjust their expectations when employees move from full-time to part-time work (Edwards & Robinson, 2000). Alternatively, managers often assume that part-timers are not committed or serious workers and underuse them (Cooper et al., 2001; Edwards & Robinson, 2000; Lewis, 2001). For example, a study of part-time police officers revealed that they were often overlooked for training and promotion (Edwards and Robinson, 2000), while part-timers in a survey of chartered accountants reported that they typically worked proportionately as many hours over their contracts as full-timers but were still regarded as not committed, and many felt that they were given less challenging assignments (Cooper et al., 2001). It is important for future research to examine the reasons why managers tend to underestimate part-time workers and what it takes to change managerial assumptions about the primacy of full-time work (Raabe, 1996, 1998). Furthermore, it is worth noting that attitudes to part-time workers appear to vary cross-nationally. A study of part-time work in the public health services in Denmark, France, and the UK suggested that while part-time work was regarded as a sign of low career commitment in the UK and France this view is less evident in Denmark where there is also more equality between male and female part-time work (Branine, 1999). Further cross-national studies may illuminate the reasons for these differences.

Managers can also undermine FWAs among full-time employees by the encouragement of long hours and face time. For example, in a case study of engineers using a combination of interviews and participant observation Perlow reveals how managers use organizational culture to control subordinates' boundaries between work and personal lives, by the various ways in which they 'cajole, encourage, coerce or otherwise influence the amount of time employees spend visibly at the workplace' (Perlow, 1998, p. 329). They do this Perlow argues by, for example, overtly valuing and rewarding long hours at the workplace and penalizing those who do not conform, setting unnecessary deadlines, and constantly monitoring employees. When commitment and productivity are difficult to quantify, as they are in most knowledge work, then they are often measured by workers' willingness to work late to meet a series of deadlines, or simply to get the work done (Lewis, 1997, 2001). Consequently managers may not only be unsupportive of FWAs but may actively undermine them.

Managers as Role Models to Subordinates and Peers

A further way in which managers can influence the outcomes of FWAs is by the ways they model work–life integration in their own lives; for example, by working reduced or flexible schedules or full-time schedules but not long

hours (Bond et al., 2002; Kossek, Barber, & Winters, 1999; Lee et al., 2000; Lewis, 2001; Raabe, 1996, 1998). Managers' working patterns send out powerful signals about what sort of working hours or schedules are acceptable. It is therefore important to understand not only what drives managers to work long hours and to expect others to do the same, but also what factors contribute to their decisions to work non-standard hours or to use FWAs in the face of what are often strong cultural and structural barriers (Raabe, 1998). A study by Kossek et al. (1999) suggests that managers' decisions about their own working patterns like those of their subordinates are influenced by perceived business impact (Kossek et al., 1999). Personal factors also play a role, with women and younger managers being more likely than other groups to say they have worked flexibly or intend to do so at some point (Kossek et al., 1999). In addition, the working patterns of their peers can exert a powerful influence on managers' behaviour. Managers in Kossek et al.'s study were much more likely to have used or to intend using flexible working practices if they had peers who had previously used them (Kossek et al., 1999). Managers who can lead the way by using FWAs themselves thus have the potential to be change agents, influencing the culture and indeed the behaviour of their peers and subordinates. However, managers often have different rules for their subordinates and themselves. For example, in a study of 'family friendly' policies and practices in 17 companies in Scotland, Bond et al. (2002) noted that, while many managers saw the importance of FWAs for their subordinates, they tended to eschew flexibility and work long hours themselves.

Research on management decision making strategies in relation to both subordinates' and their own working schedules is a promising avenue of investigation, as yet in its infancy. While policy capturing studies using hypothetical scenarios are useful for testing specific theories, more research is now needed on the factors that influence managers' decision making and behaviours in a range of real-life organizational settings. Research needs to take account of a wider range of factors influencing management assumptions and decision-making. For example, research on the impact of management training or guidelines on managing flexible workers and of feedback on colleagues' experiences of managing flexible workers would help to inform training policy. The extent to which managers are themselves under pressure, with constant deadlines, may also affect their ability to consider longer term impacts of decisions and expectations concerning subordinates' working schedules.

CONCLUSIONS

The three strands of research discussed here obviously overlap. Nevertheless, there remains a need for more connections to be made between them. Further

links could be made between research into the factors that influence organizational adoption of policies and that into management decision making relating to the day-to-day practice of managing flexible workers. For example, how do factors such as organizational size and sector, which are associated with the adoption of FWAs, impact on the way flexible work is managed in everyday practice? Research evaluating the outcomes of FWAs points to the paramount importance of organizational culture and working practices and to the need to distinguish between policy and practice. It is time for this to be recognized in all research on FWAs so that not just the adoption of policies but also the actual take-up and practice of FWAs becomes the focus of enquiry. The creation of organizational cultures that support truly flexible working arrangements to meet the needs of employees and employers may be one of the major challenges facing organizations at a time when human resources are so crucial in the global economy.

REFERENCES

Ahrentzen, S. (1990). Managing conflict by managing boundaries. How professional homeworkers cope with multiple roles at home. *Environment and Behavior*, **22**, 723–752.

Allen, T. D. (2001). Family-supportive work environments: The role of organizational perceptions. *Journal of Vocational Behaviour*, **58**, 414–435.

Auerbach, J. D. (1990). Employer-supported child care as a women-responsive policy. *Journal of Family Issues*, **11**(4), pp. 384–400.

Bailyn, L. (1993). *Breaking the Mold. Women, Men and Time in the New Corporate World*. New York: Free Press.

Bailyn, L., Rapoport, R., Kolb, D., & Fletcher, J. (1996). *Relinking Work and Family. A Catalyst for Organizational Change*. New York: Ford Foundation.

Ball, J. E. (1997). Shifting the control: Evaluation of a self scheduling flexitime rostering system. *European Nurse*, **2**(2), 100–110.

Baltes, B. B., Briggs, T. E, Huff, J. W, Wright, J. A., & Neuman, G. A. (1999). Flexible and compressed workweek schedules: A meta-analysis of their effects on work-related criteria. *Journal of Applied Psychology*, **84**(4), 496–513.

Bardoel, E. A., Tharenou, P., & Moss, S. A. (1998). Organizational predictors of work–family practices. *Asia Pacific Journal of Human Resources*, **36**(3), 31–50.

Barham, L. J., Gottlieb, B. H., & Kelloway, E. K. (1998). Variables affecting managers' willingness to grant alternative work arrangement. *Journal of Social Psychology*, **138**, 291–302.

Barnett, R. C., Gareis, K. C., & Brennan, R. T. (1999). Fit as a mediator of the relationship between work hours and burnout. *Journal of Occupational Health Psychology*, **4**(4), 307–317.

Barnett, R. C., & Hall, D. T. (2001). How to use reduced hours to win the war for talent. *Organizational Dynamics*, **29**(3), 192–210.

Bartol, K. M., & Martin, D. C. (1989). Effects of dependence, dependency threats and pay secrecy on managerial pay allocations. *Journal of Applied Psychology*, **74**, 105–113.

Bevan, S., Dench, S., Tamkin, P., & Cummings, J. (1999). *Family Friendly Employment. The Business Case* (Research Report No. 136). London: Department for Education and Employment.

Bond, S., Hyman, J., Summers, J. & Wise, S. (2002). *Family Friendly Working? Putting Policy into Practice*. York, UK: Joseph Rowntree Foundation.

Branine, M. (1999). Part time work in the public health service of Denmark, France and the UK. *The International Journal of Human Resource Management*, **10**(3), 411–429.

Brannen, J., Lewis, S., Nielson, A., & Smithson, J. (2002). *Young Europeans, Work and Family. Futures in Transition*. London: Routledge.

Brewster, C., Hegwisch, A., Lockhart, L., & Mayne L. (1993). *Issues in People Management: Flexible Working Patterns in Europe* (No. 6) London: Institute of Personnel & Development.

Burke, R. J., & Greenglass, E. R. (2001). Hospital restructuring stressors, work-family concerns and psychological well being among nursing staff. *Community, Work and Family*, **4**(1), 49–62.

Cooper, C. L., Lewis, S., Smithson, J., & Dyer, J. (2001). *Flexible Futures: Flexible Working and Work–Life Integration* (Report on Phase One). London: Institute of Chartered Accountants in England and Wales.

Dalton, D. R., & Mesch, D. J. (1990). The impact of flexible scheduling on employee attendance and turnover. *Administrative Science Quarterly*, **35**, 370–387.

Den Dulk, L. (2001). *Work-Family Arrangements in Organizations: A Cross-national Study in the Netherlands, Italy, the United Kingdom and Sweden*. Amsterdam: Rozenberg.

Den Dulk, L., & Lewis, S. (2000). Factors influencing work–family arrangements in the UK and the Netherlands. Paper presented at BPWF, University of California and Sloan Foundation Conference 'Work and Family. Expanding the Horizons', San Francisco.

Dex, S., & Schreibl, F. (2001). Flexible and family friendly working arrangements in SMEs: Business Case 2001. *British Journal of Industrial Relations*, **38**(3), 411–431.

DfEE (2000). *Work–Life Balance. Changing Patterns in a Changing World*. London: Department for Education and Employment.

Dubrin, A. (1991). Comparison of the job satisfaction and productivity of telecommuters versus in-house employees. *Psychological Reports*, **68**, 1223–1224.

Edelman, L. B. (1990). Legal environments and organizational governance: The expansion of due process in the American workplace. *American Journal of Sociology*, **95**(6), 1401–1440.

Edwards, C., & Robinson, O. (2000). Managing part timers in the police service: A study of flexibility. *Human Resource Management Journal*, **9**(4) 5–18.

Evans, J. E. (2000). Firms' contribution to the reconciliation between work and family life: Experiences in OECD countries. Paper given at the *Seminar on Family-Friendly Employment* (organized by the Family Policies Studies Centre, London, November 2000).

Fletcher, J., & Rapoport, R. (1996). Work–Family issues as a catalyst for organizational change. In S. Lewis, & J. Lewis (eds), *The Work–Family Challenge* (pp. 142–158). London: Sage.

Forth, J., Lissenburgh, S., Callender, C., & Millward, N. (1997). *Family Friendly Working Arrangements in Britain, 1996* (Research Report No. 16). London: Policy Studies Institute, Department for Education and Employment.

Fried, M. (1998). *Taking Time. Parental Leave Policy and Corporate Culture*. Philadelphia: Temple University Press.

Friedman, D. E., & Johnson, A. A. (1996). Moving from programs to culture change: The next stage for the corporate work–family agenda. In S. Parasuraman, & J. H. Greenhaus (eds), *Integrating Work and Family. Challenges and Choices for a Changing World*. Westport, CT: Quorum.

Friedman, S. F., & Greenhaus, J. H. (2000). *Work and Family—Allies or Enemies*. New York: Oxford University Press.

Frolick, M. N., Wilkes, R. B., & Urwiler, R. (1993). Telecommuting as a workplace alternative: An identification of significant factors at home in American firms' determination of work-at-home policies. *Journal of Strategic Information Systems*, **2**, 206–222.

Frone, M., Russell, M., & Cooper, M. (1992). Antecedents and outcomes of work–family conflict: Testing a model of the work–family interface. *Journal of Applied Psychology*, **77**, 65–75.

Frone, M., Yardley, J., & Markel, K. (1997). Developing and testing an integrative model of the work–family interface. *Journal of Vocational Behavior*, **50**, 145–167.

Galinsky, E., Friedman, D. E., & Hernandez, C. A. (1991). *The Corporate Reference Guide to Work–Family Programs*. New York: Families and Work Institute.

Galinsky, E., & Johnson, A. A. (1998). *Reframing the Business Case for Work–Life Initiatives*. New York: Families and Work Institute.

Glass, J., & Fujimoto, T. (1995). Employers characteristics and the provision of family responsive policies. *Work and Occupation*, **22**(4), 380–411.

Glass, J. L., & Estes, S. B. (1997) The family responsive workplace. *Annual Review of Sociology*, **23**, 289–313.

Goff, S. J., Mount, M. K., & Jamison, R. L. (1990). Employer supported child care, work/family conflict, and absenteeism: A field study. *Personnel Psychology*, **43**, 793–809.

Golden, L. (2001). Flexible work schedules. Which workers get them? *American Behavioral Scientist*, **44**(7), 1157–1178.

Goodstein, J. D. (1994). Institutional pressures and strategic responsiveness: Employer involvement in work–family issues. *Academy of Management Journal*, **37**(2), 350–382.

Greenberg, J. (1996). *The Quest for Justice in the Workplace*. Thousand Oaks, CA: Sage.

Greenhaus, J. H., & Parasuraman, S. (1999). Research on work, family and gender: Current status and future directions. In G. Powell (ed.), *Handbook of Gender and Work*. Thousand Oaks, CA: Sage.

Grover, S. L. (1991). Predicting the perceived fairness of parental leave policies. *Journal of Applied Psychology*, **76**, 247–255.

Grover, S. L., & Crooker, K. J. (1995). Who appreciates family-responsive human resource policies: The impact of family-friendly policies on the organizational attachment of parents and non-parents. *Personnel Psychology*, **48**, 271–288.

Haddon, L. (1992). *Clerical Teleworking and Family Life*. Martlesham, UK: British Telecom.

Harker, L. (1996). The family friendly employer in Europe. In S. Lewis, & J. Lewis (eds), *The Work Family Challenge. Rethinking Employment*. London: Sage.

Harker, L., & Lewis, S. (2001). Work–life policies: Where should the government go next? In N. Birkett (ed.), *A Life's Work: Achieving Full and Fulfilling Employment*. London: Institute of Public Policy Research.

Hill, E. J., Hawkins, A. J., Ferris, M., & Weitzman, M. (2001). Finding an extra day a week: The positive influence of perceived job flexibility and work and family life balance. *Family Relations*, **50**(1), 49–58.

Hill, E. J., Hawkins, A. J., & Miller, B. C. (1996). Work and family in the virtual office. Perceived influences of mobile telework. *Family Relations*, **45**(3), 293–301.

Hochschild, A. (1997). *The Time Bind. When Work Becomes Home and Home Becomes Work*. New York: Henry Holt.

Hogarth, T., Hasluck, C., Pierre, G., Winterbotham, M., & Vivian, D. (2000). *Work–Life Balance 2000: Baseline Study of Work–Life Balance Practices in Great Britain*. London: Department for Education and Employment.

Hohl, K. (1996). The effects of flexible working arrangements. *Nonprofit Management and Leadership*, **7**, 69–86.

Holt, H., & Thaulow, I. (1996). Formal and informal flexibility in the workplace. In S. Lewis, & J. Lewis (eds), *The Work Family Challenge. Rethinking Employment*. London: Sage.

Ingram, P., & Simons, T. (1995). Institutional and resource dependence determinants of responsiveness to work–family issues. *Academy of Management Journal*, **38**(5), 1466–1482.

Kelloway, E. K., Gottlieb, B. H., & Barham, L. (1999). The source, nature and direction of work–family conflict: A longitudinal investigation. *Journal of Occupational Health Psychology*, **4**(4), 337–346.

Kivimaki, R. (1998). How work is structured by the family? The impacts of parenthood on the work community. Paper given at the *Gender, Work and Organisation Conference, Manchester*.

Klein, K. J., Berman, L. M., & Dickson, M. W. (2000). May I work part-time? An exploration of predicted employer responses to employee request for part time work. *Journal of Vocational Behavior*, **57**, 85–101.

Kogi, K., & Martino, V. G. (1995). Trends in the participatory process of changing shiftwork arrangements. *Work and Stress*, **9**, 298–304.

Kossek, E., Barber, A., & Winters, D. (1999). Using flexible schedules in the managerial world: The power of peers. *Human Resource Management*, **38**(1), 33–46.

Kossek, E., & Ozeki, C. (1998). Work–family conflict policies and the job–life satisfaction relationship. *Journal of Applied Psychology*, **83**, 139–149.

Kossek, E. E., Dass, P., & DeMarr, B. (1994). The dominant logic of employer-sponsored work and family initiatives: Human resource managers' institutional role. *Human Relations*, **47**(1), 1121–1149.

Kossek, E. E., Nichol, V. (1992). The effects of sire childcare on employee attitudes and performance. *Personnel Psychology*, **45**, 485–509.

Kossek, E. E., & Ozeki, C. (1999). Bridging the work–family policy and productivity gap: A literature review. *Community, Work and Family*, **2**(1), 7–32.

Krausz, M., Sagie, A., & Biderman, Y. (2000). Actual and preferred work scheduling control as determinants of job related attitudes. *Journal of Vocational Behavior*, **56**, 1–11.

Lee, M. D., McDermid, S. M., & Buck, M. L. (2000). Organizational paradigms of reduced-load work: Accommodations, elaboration and transformation. *Academy of Management Journal*, **43**(6), 1211–1226.

Lewis, S. (1997). Family friendly organizational policies: A route to organizational change or playing about at the margins. *Gender, Work and Organisation*, **4**, 13–23.

Lewis, S. (2000). Organisational change and gender equity. Case studies from the UK. In L. Haas, P. Hwang, & G. Russell (eds), *Organisational Change and Gender Equity*. London: Sage.

Lewis, S. (2001). Restructuring workplace cultures: The ultimate work–family challenge? *Women in Management Review*, **16**(1), 21–29.

Lewis, S., & Cooper, C. (1995). Balancing the work family interface. A European perspective. *Human Resource Management Review*, **5**, 289–305.

Lewis, S., Cooper, C. L. (1999). The work–family research agenda in changing contexts. *Journal of Occupational Health Psychology*, **4**, 4382–4393.

Lewis, S., Kagan, C., & Heaton, P. (1999). Economic and psychological benefits from employment. The experiences of mothers of disabled children. *Disability and Society*, **14**(4), 561–575.

Lewis, S., & Lewis, J. (1997). Work, family and well being. Can the law help? *Legal and Criminological Psychology*, **2**, 155–167.

Lewis, S., & Smithson, J. (2001). Sense of entitlement to support for the reconciliation of employment and family life. *Human Relations*, **54**(11), 1455–1483.

Lewis, S., Smithson, J., Brannen, J., Das Dores Guerreiro, M., Kugelberg, C., Nilsen, A., & O'Connor, P. (1998). *Futures on Hold. Young Europeans Talk about Combining Work and Family*. London: Work–Life Research Centre.

Lewis, S., Smithson, J., Cooper, C. L., & Dyer, J. (2002). *Flexible Futures: Flexible Working and Work–Life Integration* (Report on Phase Two). London: Institute of Chartered Accountants in England and Wales.

Lewis, S., Taylor, K. (1996). Evaluating the impact of employer family friendly policies. A case study. In S. Lewis, & J. Lewis (eds), *The Work Family Challenge* (pp. 112–127). London: Sage.

Lyness, K. S., Thompson, C. A., Francesco, A. M., & Jusiesch, M. K. (1999). Work and pregnancy: Individual and organizational factors influencing organizational commitment, timing of maternity leave and return to work. *Sex Roles*, **7/8**(41), 485–507.

MacDermid, S., Litchfield, L., & Pitt-Catsouphes, M. (1999). Organisational size and work–family issues. *Annals of the American Academy of Political and Social Science*, **562**, 111–126.

Martens, M. F. J., Nijhuis, F. J. N, van Boxtel, M. P. J., & Knottnerus, J. A. (1999). Flexible work schedules and mental and physical health. A study of a working population with non-traditional working hours. *Journal of Organizational Behaviour*, **22**(1), 35–47.

Maume, D. J., & Bellas, M. L. (2001). The overworked American or the time bind? *American Behavioral Scientist*, **44**(7), 1137–1156.

Maume, D., & Houston, P. (2001) Job segregation and gender differences in work–family spillover among white collar workers. *Journal of Family and Economic Issues*, **22**(2), 171–189.

Milliken, F. J., Martins, L. L., & Morgan, H. (1998). Explaining organizational responsiveness to work–family issues: The role of human resource executives as issue interpreters. *Academy of Management Journal*, **41**(5), 580–592.

Morgan, H., & Milliken, F. J. (1992). Keys to action: Understanding differences in organizations' responsiveness to work and family issues. *Human Resource Management*, **31**(3), 227–248.

Moss, P., & Deven, F. (1999). *Parental Leave: Progress or Pitfall?* Leuven, Belgium: Centrum voor Bevolkings en Gezingstudie.

Netemeyer, R., Boles, J., & McMurrian, R. (1996). Development and validation of work–family conflict and family–work conflict scales. *Journal of Applied Psychology*, **81**, 400–410.

Olsen, M. H. (1987). Telework: Practical experience and future prospects. In R. E. Kraut (ed.), *Technology and the Transformation of White Collar Work*. Hillsdale, NJ: Lawrence Erlbaum.

Osterman, P. (1995). Work–family programs and the employment relationship. *Administrative Science Quarterly*, **40**, 681–700.

Parker, L., & Allen, T. D. (2000). Work/Family benefits: Variables related to employees' fairness perceptions. *Journal of Vocational Behavior*, **58**, 453–468.

Perlow, L. A. (1998). Boundary control: The social ordering of work and family time in a high tech organisation. *Administrative Science Quarterly*, **43**, 328–357.

Powell, G. N., & Mainiero, L. A. (1999). Managerial decision making regarding alternative work arrangements. *Journal of Occupational and Organizational Psychology*. **72**(1), 41–57.

Presser, H. B. (1998) Towards a 24 hour economy. The US experience and implications for the family. In D. Vannoy and P. Dubeck (eds), *Challenges for Work and Family in the Twenty First Century*, New York: Aldine De Gruyter.

Prutchno, R., Litchfield, L., & Fried, M. (2000). *Measuring the Impact of Workplace Flexibility*. Boston: Boston College Center for Work and Family.

Purcell, K., Hogarth, T., & Simm, C. (1999). *Whose Flexibility? The Costs and Benefits of Non-standard Working Arangements and Contractual Relations*. York, UK: Joseph Rowntree Foundation.

Raabe, P. (1996). Constructing pluralistic work and career arrangements. In S. Lewis, & J. Lewis (eds), *The Work–Family Challenge. Rethinking Employment*. London: Sage.

Raabe, P. H. (1998). Being a part time manager. One way to combine family and career. In D. Vannoy, & P. Dubeck (eds), *Challenges for Work and Family in the Twenty First Century*, New York: Aldine De Gruyter.

Rapoport, R., Bailyn, L., Fletcher, J., & Pruitt, B. (2002). *Beyond Work–Family Balance. Advancing Gender Equity and Workplace Performance*. San Francisco: Jossey-Bass.

Ralston, D. A. (1989). The benefits of flexitime. Real or imaginary? *Journal of Organizational Behaviour*, **10**, 369–373.

Rowe, B. R., & Bentley, M. T. (1992). The impact of family on home based work. *Journal of Family and Economic Issues*, **13**, 279–297.

Schreibl, F., & Dex, S. (1998). Should we have more family friendly policies? *European Management Journal*, **16**(5), 586–599.

Smith, P. A., & Wedderburn, A. A. (1998). Flexibility and long shifts. *Employee Relations*, **20**(5), 481–489.

Standen, P., Daniels, K., & Lamond, D. (1999). The home as a workplace: Work–Family interaction and psychological well-being in telework. *Journal of Occupational Health Psychology*, **4**(4), 368–381.

Stanworth, C. M. (1999). A best case scenario? Non manual part time work and job sharing in UK local government in the 1990s. *Community, Work and Family*, **2**(3), 295–310.

Steward, B. (2000). Changing times: The meaning, measurement and time in teleworking. *Time and Society*, **9**(1), 57–74.

Sullivan, C. (2000). Space and the intersection of work and family in homeworking households. *Community, Work and Family*, **3**(2), 185–204.

Sullivan, C., & Lewis, S. (2001). Homebased teleworking, gender and the synchronisation of work and family. *Gender, Work and Organisation*, **8**(2), 123–145.

Tausig, M., & Fenwick, R. (2001). Unbinding time: Alternative work schedules and work–life balance. *Journal of Family and Economic Issues*, **22**(2), 101–119.

Thompson, C., Beauvais, L., & Lyness, K. (1999). When work–family benefits are not enough. The influence of work–family culture on benefit utilization, organizational attachment and work–family conflict. *Journal of Vocational Behaviour*, **54**, 392–415.

Thomas, L., & Ganster, D. (1995). Impact of family supportive work variables on work family conflict and strain: A control perspective. *Journal of Applied Psychology*, **80**, 6–15.
Wood, S. (1999). Family friendly management: Testing the various perspectives. *National Institute Economic Review*, **168**, 99–116.
Wood, S., DeMenezes, L., & Lasaosa, A. (2003). Family-friendly management in the UK: Testing various perspectives. *Industrial Relations* (in press).
Young, M. (1999). Work–family backlash. Begging the question, what's fair? *Annals of the American Academy of Political Science*, **562**, 32–46.

Chapter 2

ECONOMIC PSYCHOLOGY

Erich Kirchler and Erik Hölzl
University of Vienna

Economic psychology is concerned with understanding human experience and human behaviour in economic contexts. Textbooks of economic psychology usually provide an introduction to the theoretical and normative fundamentals of human behaviour and the anomalies of everyday decision making. Further topics are economic socialization and lay theories, consumer markets from the perspective of households and businesses, and labour markets. Additional areas on the national level are poverty and affluence, money and the psychology of inflation, taxation behaviour, housework, and the shadow economy.

The present review first gives an overview on the diverse research areas in economic psychology by reporting an analysis of articles published in the *Journal of Economic Psychology*, from its inception in 1981 to 2001. Since the field is influenced by two scientific disciplines, a short outline of the history of economic psychology is given to provide a background for understanding the sometimes conflicting perspectives of psychology and economics. The chapter then focuses on decision making behaviour and topics in economic psychology that featured prominently in the *Journal of Economic Psychology* during the last six years, from 1996 to 2001. In addition, the review considers the standard works in the field, and, where necessary for understanding, other publications.

RESEARCH AREAS AND PERSPECTIVES

In the period up to the end of 2001, over 650 articles (apart from book reviews, short commentaries, and the like) appeared in the *Journal of Economic Psychology*. From the beginning of 1996 alone, the number of articles

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

totalled 224. In order to establish the topic areas covered, Schuldner (2001) analysed and categorized the content of the titles, keywords, and abstracts. Categorization of the publications proved difficult; an attempt to achieve this using the keywords given in the articles was abandoned as impossible because certain concepts were either heterogeneous or missing altogether. The keywords used in the PsycINFO database also proved unsatisfactory and often misleading. Instead, a step-by-step, inductive set of categories was constructed with the aid of five colleagues in the field. In addition, the topic areas of two textbooks (Kirchler, 1999; Lea, Tarpy, & Webley, 1987) were used in structuring the content categories. Once convergence had been achieved and the articles had been satisfactorily and unambiguously assigned, the category system could be accepted. Table 2.1 lists content categories and frequencies of publications.

Approximately two-thirds of the publications in the *Journal of Economic Psychology* relate to topics in economic psychology (65%). Market and consumer psychology feature strongly with 29%. A third topic area is environmental psychology (5%). In economic psychology, the focus is on research into decision making (19%), with studies of choice and decision making by individuals (11%), and in social interactions, frequently from the perspective of game theory (7%). Topics relating to financial behaviour included investment decisions, savings behaviour, debts and credits in private households (5%), and financial markets (4%). Considerable space was also devoted to taxation behaviour (7%), the labour market (7%), economic socialization and lay theories (5%). Although the *Journal of Economic Psychology* does treat political aspects of economics such as economic growth and welfare, tax policies, and reforms (5%), these topics are represented particularly prominently in the *Journal of Socio-Economics*, whose target group consists mainly of economists interested in behavioural science. In the last six years, to which the present review relates, interest in decision making and choice has markedly increased (from 15% in 1981 to 1995 up to 26% in the years from 1996 to 2001). Not surprisingly, work on money and on the transition from national European currencies to the euro increased (from 2% to 6%). Studies on inflation have decreased (from 4% to 0%), as have those on consumer behaviour (from 31% to 25%) and environmental psychology (from 7% to 3%).

The above analysis of topics covered in the *Journal of Economic Psychology* casts light on the content of research. In addition, an analysis of authors and publications cited permits an identification of the main perspectives from which those topics are being studied. Schuldner (2001) found over 12,000 quotations from the literature in the period 1991 to the beginning of 2001. These covered 8,031 different studies, books, and journals. Previous publications in the *Journal of Economic Psychology* (4.3%) were most often cited in these articles, followed by publications in the *Journal of Consumer Research* (3.9%), *Journal of Personality and Social Psychology* (2.3%), and *American*

Table 2.1 Publications in the *Journal of Economic Psychology* from 1981 to 2001.

Categories	Total	1981–85	1986–90	1991–95	1996–01
Economic psychology	*424*	*61*	*82*	*121*	*160*
Theory and history (e.g., theoretical frameworks, life and work of scientist)	20	5	2	7	6
Choice and decision theory	122	14	15	33	60
Decision theory (e.g., decision-making under risk, choice behaviour, preference formation)	74	10	10	20	34
Co-operation/game theory	48	4	5	13	26
Socialization (e.g., lay theories, economic socialization)	34	4	17	6	7
Firm (e.g., firm behaviour, entrepreneurship)	18	1	1	7	9
Labour market (e.g., labour supply, work experiences, income and wage, unemployment)	44	8	6	15	15
Marketplace (e.g., pricing, price competition)	11	1	1	4	5
Financial attitudes and behaviour	59	3	9	22	25
Household financial behaviour (e.g., saving, credit and loan, debts)	35				
Investment/stock market	24				
Money (e.g., money in general, euro)	20	1	2	4	13
Inflation	17	5	10	1	1
Tax (e.g., tax attitudes, evasion)	44	10	12	12	10
Government and policy (e.g., welfare, growth, and prosperity)	35	9	7	10	9
Consumer psychology	*188*	*32*	*47*	*52*	*57*
Consumer attitudes	37				
Consumer behaviour	86				
Consumer expectations	46				
Marketing and advertisement	19				
Environmental psychology and other issues	*39*	*24*	*2*	*6*	*7*
Total	*651*	*117*	*131*	*179*	*224*

Note: The frequencies given for the principal categories (economic psychology, consumer psychology, and environmental psychology) incorporate in each case the figures for the subsidiary categories.

Economic Review (2.0%). The main subject areas of these journals indicate that economic psychology is mainly pursued from the perspective of social psychology and economics, but that there is also much space given to consumer research. The most cited authors were Daniel Kahneman, Nobel Laureate in 2002, and Amos Tversky, as well as Richard Thaler. All three have become well known for their direction-setting contributions to decision-making research and publications in *Econometrica*, *Science*,

American Psychologist, *Journal of Economic Behavior and Organization*, *Marketing Science*, etc. (see the obituary for Amos Tversky by van Raaij, 1998). Werner Güth's contributions in the *Journal of Economic Behavior and Organization* and the *Journal of Economic Psychology* were frequently quoted in articles on game theory. Studies on consumer psychology and on the symbolism of goods drew particularly on the work of Russel W. Belk and Helga Dittmar published in the *Journal of Consumer Research*, the *British Journal of Social Psychology*, and in monographs. George Katona's *Psychological Economics* from 1975 and the 1987 textbook *The Individual in the Economy* by Stephen E. G. Lea, Roger M. Tarpy, and Paul Webley are frequently quoted. The articles published in 1989 and 1990 in the *Journal of Economic Psychology* by Karl-Erik Wärneryd and Fred van Raaij also feature prominently.

THE HISTORY OF ECONOMIC PSYCHOLOGY

The economic sciences study decisions on the use of scarce resources for the purpose of satisfying a multiplicity of human needs. People normally find themselves unable to satisfy all their needs, and are forced to choose between alternatives; their choice of one option in turn involves the pain of renouncing the advantages of all the other options. In economics, decisions on the allocation of scarce resources are described on the premise of rationality and maximization of utility. Economics has constructed complex, formal, decision-making models to explain and predict economic behaviour, starting from only a small number of axioms on the logic of human behaviour. These models often do not consider psychology.

Classical economics, which traces its origins to Adam Smith's (1776) *Wealth of Nations*, found itself challenged toward the end of the 19th century. Thorstein Veblen (1899) opposed the basic assumptions of rationality regarding decision-making goals and utility maximization with his findings on conspicuous consumption, showing that some goods become particularly desirable when the price rises. He expressed the criticism that economics does not consider cultural factors and social change. Wesley C. Mitchell (1914, p. 1) introduced his work on human behaviour and economics with the observation, 'A slight but significant change seems to be taking place in the attitude of economic theorists toward psychology', and closed with the prediction, '... economics will assume a new character. It will cease to be a system of pecuniary logic, a mechanical study of static equilibria under non-existent conditions, and become a science of human behavior' (p. 47). Clark (1918, p. 4) wrote: 'The economist may attempt to ignore psychology, but it is a sheer impossibility for him to ignore human nature, for his science is a science of human behavior. Any conception of human nature that he may adopt is a matter of psychology, and any conception of

human behavior that he may adopt involves psychological assumptions, whether these be explicit or not.'

Economics and psychology showed an interest in the other discipline early on. It has long been beyond dispute on both sides that psychology and economics, and likewise sociology and economics (Swedberg, 1991), have not only extensive common boundaries, but also an overlap in the questions they pose. The main argument was that neither money, inflation rate, nor unemployment figures by themselves influence each other, but that people act and interact in a given economic environment and thereby change it. Economic data are the aggregated measurements of individual behaviour; in other words, economics consists for the most part of aggregated psychology. However, warnings and advice of representatives interested in interdisciplinary approaches found only limited reception among the majority of their 'orthodox' colleagues.

Gabriel Tarde (1902) in France was probably the first to use the term 'economic psychology'. His *La Psychologie Économique* drew attention to the need to analyse economic behaviour from a psychological perspective. He particularly criticized Adam Smith for not having incorporated his knowledge of human psychology, which he had demonstrated in his writings, into his concept of the economy. The existence of ventures in psychological thinking in Smith's work is described by Khalil (1996) in an essay on the 'Theory of moral sentiments'. Smith did not only emphasize the satisfaction of pecuniary, constitutive utility. Contrary to the orientation of modern welfare economics, he also stressed that self-respect is a chief ingredient of satisfaction. Hugo Münsterberg (1912) is seen as the initiator of this field of thought in the German-speaking world. He emphasized the need for close co-operation between psychology and the economic sciences. He began with studies on sociotechnology, on monotony in working life, on the selection of staff, and experimental research into the effects of advertising. However, his comprehensive approach was then put in the shade by developments in occupational and organizational psychology.

In the late 1940s, George Katona and Günther Schmölders began to design a psychology of macroeconomic processes. Katona's (1951, p. 9f) view of economic psychology is clearly described in the following quotation: '. . . the basic need for psychology in economic research consists in the need to discover and analyze the forces behind economic processes, the forces responsible for economic actions, decisions and choices . . . Economics without psychology has not succeeded in explaining important economic processes and "psychology without economics" has no chance of explaining some of the most common aspects of human behavior.'

Together with Burkhard Strümpel, Katona criticized the economic model of the time as limited: 'The savings rate, for example, is seen as dependent on total income, the price level as a function of the money supply, the level of demand as determined by prices. The human being as the active agent at the

centre of this dynamic picture is airbrushed out as an anonymous "black box" ... In fact, however, the human being who occupies the position in between his environment and the economic outcome of his behaviour is full of self-will. He is dominated by prejudices, mood-driven, impulsive, and poorly informed. He is exposed to changing influences, but forgets or neglects much of the fruit of experience, even occasionally jettisoning principles and overall concepts of the world. He transfers experiences and the wisdom they bring from one sphere of life to another, and even manages to alter economic expectations when important non-economic events occur. He learns.' (Strümpel & Katona, 1983, p. 225).

Among economists, the voice of Herbert Simon attracted particular attention. He saw restrictions to the validity of the widely accepted rational model, especially in man's limited cognitive capacities (for a collection of articles by Simon and scientists in the field see Earl, 2001). In an obituary for Herbert Simon, Augier (2001) quotes some important statements that express Simon's position clearly: 'The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world—or even for a reasonable approximation to such objective rationality' (Simon, 1982, p. 204). 'For the first consequence of the principle of bounded rationality is that the intended rationality of an actor requires him to construct a simplified model of the real situation in order to deal with it. He behaves rationally with respect to this model, and such behavior is not even approximately optimal with respect to the real world. To predict his behavior, we must understand the way in which this simplified model is constructed, and its construction will certainly be related to his psychological properties as a perceiving, thinking, and learning animal' (Simon, 1957, p. 199).

The development of economic models on the basis of a small number of axioms of human behaviour and the preoccupation with economic indices that rested on a set of norms of human behaviour, the realism of which was questioned less and less as the formulaic language of mathematical logic became more and more attractive, caused unease among economists. It stimulated an interest in everyday behaviour in the economic context. Economic psychology sets descriptive economic models against normative ones. In their book about the social psychology of economic behaviour, Furnham and Lewis (1986) make the useful distinction between economic psychology and psychological or behavioural economics. On the one hand, psychologists try to understand experience and behaviour in the economic context, and, on the other hand, economists who have found the straitjacket of traditional theoretical principles too restricting adopt findings from scientific psychology into the formal models.

'The ultimate criterion of all economic activities and economic policy is human well-being,' wrote van Veldhoven (1988, p. 53). He stated: 'In the

end, all distribution of scarce means and goods serves the fulfilment of needs and aspirations and the achievement of satisfaction of individuals and groups. Thus, economic reality cannot adequately be understood without the analysis of the subjective and psychological dynamics that underlie and guide economic behaviour of both individuals and groups.' From about the 1970s onward, social scientists and economists have emphasized the importance of economic psychology and behavioural economy (Wärneryd, 1988, 1993). Lunt (1996), however, criticizes the attempts of economically oriented psychologists who adopt economists' agendas and introduce psychological insight into elaborate economic models. He claims that psychologists should start to examine economic theory to open new lines of collaboration that will allow them to apply their own conception of psychology to economics.

The International Association for Research in Economic Psychology (IAREP) was founded primarily by European psychologists and economists, and has the influential *Journal of Economic Psychology* since 1981. The Association successfully bridges psychology and economics. In the USA, there are two related associations consisting for the most part of a combination of economists and sociologists, concerned with behavioural concepts in economic matters: the Society for the Advancement of Socio-Economics (SASE) and the Society for the Advancement of Behavioral Economics (SABE), which publishes the *Journal of Socio-Economics*. Apart from the journals, economic psychology is represented in a number of introductory works (Antonides, 1991; Earl & Kemp, 1999; Ferrari & Romano, 1999; Furnham & Lewis, 1986; Kirchler, 1999; Lea et al., 1987; van Raaij, van Veldhoven, & Wärneryd, 1988; Webley, Burgoyne, Lea, & Young, 2001; Wiswede, 2000).

DECISION MAKING: UTILITY MAXIMIZATION AND RATIONALITY

Economic management means making decisions. The assumptions behind the metaphor of 'Homo oeconomicus' are that individuals make rational decisions, and that the option chosen from a set of alternatives is the best for the person concerned. People's decisions and behaviour are governed by the rules of logic. A small number of axioms form the basis for complex models of optimal decision making on the part of individuals and groups engaged in managing their economic affairs (Gravelle & Rees, 1981):

(a) When the best of a bundle of alternatives is to be chosen, an individual must clarify the characteristics of the various alternatives. These characteristics must be assessed, and all the apparently available options compared with each other. According to the principle of completeness, individuals are able to place alternatives in order of preference. In

other words, they establish relationships between the alternatives, such that alternative A is better than or equal to alternative $B(A \succeq B)$, or B as good as or preferred to $A(A \preceq B)$, or the individual is indifferent between A and B $(A \approx B)$.

(b) According to the principle of transitivity, individuals create consistent orders of preference, and do not change their preferences arbitrarily. If, for example, a consumer believes alternative A to be better than or as good as alternative B, which is in turn better than or as good as alternative C, this consumer must also believe that A is better than or as good as C (if $A \succeq B$ and $B \succeq C$, then $A \succeq C$). If alternative A is as good as B and B as good as C, then the individual must also be indifferent between A and C (if $A \approx B$ and $B \approx C$, then $A \approx C$). This means that an alternative can belong to only one indifference set.

(c) The principle of reflexivity postulates that every bundle of alternatives is as good as itself $(A \approx A)$.

(d) Gravelle and Rees (1981) quote non-satiation as a further basic assumption. According to this principle, one bundle of alternatives will be preferred to another if it contains more of at least one good, and has the same quantity of other goods as the other.

(e) The fifth assumption, the axiom of continuity, states that it is possible to compensate for the loss of a certain quantity of good A by a certain quantity of good B, so that a person is indifferent to the quantity combinations (A, B) and $(A - X, B + Y)$.

(f) Lastly, the assumption is made that, when individuals possess a small quantity of good A and a large quantity of good B, they will only be indifferent to the loss of part of good A if they receive in addition a comparatively large quantity of good B. This is the axiom of convexity, and conforms to the law of satiation, according to which the relative increase in utility by one additional unit of a good diminishes with the availability of that good.

Utility maximization (frequently for egoistic goals, but sometimes for altruistic ones) and rationality are the fundamental assumptions of economics. Based on these assumptions, economics makes predictions of human behaviour in typical economic contexts, but also in other contexts, such as romantic relationships or criminality. The neoclassical paradigm has also inspired various avenues in psychology, such as theories of interaction between people in public settings and in intimate relationships (e.g., social exchange theories). These have been celebrated as ‘deromanticized’ universal theories, but also condemned as technical elaborations removed from reality.

The assumptions of rational theory or of Homo oeconomicus were often criticized, sometimes, however, on distorted grounds. Even economists do not exclusively view the human being as ‘purposeful–rational, following pure

considerations of utility, utterly subject to striving for gain, and equipped with the capacity to adapt to changing constellations of the market on the basis of a complete knowledge of market data (conditions of supply and demand), a state, therefore, of being totally informed (market transparency), with unbounded speed of reaction in adapting to changes in constellations of the market and acting accordingly, aiming at the greatest achievable utility' (von Rosenstiel & Ewald, 1979, p. 19). However, humanity is also not seen as drive-oriented, of limited cognitive ability, and thus often inconsistent by nature. The question that unsettles the very foundations of economics is whether human beings actually do pursue their goals in an economically logical way. What is it that people wish to or indeed can maximize? Is it egoistic profit, for themselves and others, or do they strive to act in accordance with society's moral demands? How consistent are orders of preference? Critics have also pointed out that in economic theory individuals are detached from their social context, and observed in isolation from other people, as if they operated in a social vacuum according to the principles of utility maximization and rationality. However, there are differences between isolated individuals who wish to act rationally, and the members of collective groups acting within the limits of rules and norms (Etzioni, 1988).

The clear formulae of economics have a certain fascination. At the same time, however, criticism aims at proceeding from the starting point of an unrealistic picture of humans, even if this picture is claimed to be a model of the average, free of unsystematically varying individual irrationality, and achieved by taking the aggregate of multiple individual actions. Frey (1990) distinguishes four possible states of individual and aggregated behaviour: individual behaviour may either correspond to or deviate from economic assumptions, and on the aggregated social level the predictions of the economic model may be fulfilled or not. The most desirable situation for followers of rational theory is when action on the individual level is rational and aimed at maximizing utility, and rational behaviour is also perceptible on the aggregated level. The second acceptable situation is where anomalies exist in individual behaviour, but disappear in the aggregation process, as happens in markets under conditions of perfect competition. However, in some cases, individual behaviour can be entirely rational, but the outcome on the collective level deviates from the rational model. This occurs when a particularly high value is placed on private goods, but public goods are devalued. Phenomena of this case are known as free riding; examples are tax evasion, offences against the environment, and the extensive use of common resources. Finally, in some cases, anomalies can be observed on both the individual and the collective level. Examples of such anomalies are those described as judgement heuristics and biases, where systematic deviations from rationality are transferred from the individual to the aggregated level of society as a whole.

DECISIONS: PSYCHO-LOGIC AND COOPERATION

The concepts of rationality and utility maximization are based on the assumption that people follow the laws of logic in choices and decision-making situations. In fact, even in situations of little complexity, the basic assumptions of economics are seen to be contravened. For a start, it is not certain whether individuals do generally try to maximize their own profit. Frijters (2000) found, for instance, only limited support for the hypothesis that people try to maximize general satisfaction with life. Pingle (1997) found that people often choose the option prescribed by the authorities rather than the one optimal to themselves.

The premise that decision-makers can rank-order the alternatives with respect to an overall ordinal value or utility, in order to choose the 'best', has frequently been shown to be violated. Li (1996) conducted an experiment in which two fundamental rational decision axioms, transitivity and independence of alternatives, were systematically contravened. Zwick, Rapoport, and Weg (2000) found violations of the invariance axiom in sequential bargaining. Huck and Weizsäcker (1999) report that subjects do not react to risk in a way that is consistent with stable expected-utility functions. When future private and social benefits are evaluated, the discounted expected-utility model, predicting exponential discounting, serves as a basic building block in modern economic theory. However, Loewenstein and Prelec (1992), Thaler (1981), and others show that subjects violate the predictions of the model in certain circumstances. Hyperbolic discounting seems to match subjects' behaviour better than exponential discounting (Laibson, 1997). The classical model also provides an inadequate explanation for the fact that people vote. The expected benefits from voting in a large-scale election are generally outweighed by the cost of the act (Downs, 1957; Schram & Sonnemans, 1996). Social norms and inter- and intra-group relations seem to provide a better explanation of voting behaviour than simple individual utility maximization (Cairns & van der Pol, 2000; Mador, Sonsino, & Benzion, 2000; Schram & van Winden, 1991).

The more complex and the less transparent a situation is, the more subjects deviate from what the rational model predicts. People behave differently according to the situation. Framing effects as described below show clearly how risk aversion can change. On the Internet, subjects bid higher for lotteries and standard deviations of bids are larger than in classrooms (Shavit, Sonsino, & Benzion, 2001). Besides the situational constraints, individual differences play a relevant role in economic behaviour. Supphellen and Nelson (2001) found effects of personality and motive structure for donating behaviour. Powell and Ansic (1997) report gender differences in risk behaviour and strategies in financial decision making. Such differences can be viewed as general traits, or as arising from context factors. The results of experiments conducted by Powell and Ansic (1997) show that women are

less risk seeking than men, irrespective of task familiarity, framing, costs, and problem ambiguity. The results also indicate that men and women adopt different strategies in financial decision environments but that these strategies have no significant impact on performance. Because strategies are more easily observed than risk preference or outcomes in everyday decisions, strategy differences may reinforce stereotypical beliefs that women are less competent financial managers. Boone, de Brabander, and van Witteloostuijn (1999) argue that co-operation in bargaining and game settings depends on personality variables. Internal locus of control, high self-monitoring, and high sensation-seeking were systematically associated with co-operative behaviour, whereas Type A behaviour was negatively correlated with co-operation.

People often fail to grasp the full range of alternatives in order to select the best, resulting partly from lack of time, and partly from limited cognitive abilities and a lack of motivation to collect and process all the relevant information. There is clear confirmation of this in decision-making situations involving risk. In addition, people do not always behave selfishly in interactions; they co-operate even when their partners could exploit the situation in a harmful way. The occurrence of reciprocity and co-operation is confirmed in particular in game theory and in experimental markets (Fehr, Gächter, & Kirchsteiger, 1997; see also, e.g., Gigerenzer, Todd, & the ABC Research Group, 1999; Güth, 2000; Jungermann, Pfister, & Fischer, 1998; Lundberg, 2000; Mellers, Schwartz, & Cooke, 1998, Wärneryd, 2001; for a collection of classical articles on anomalies in decision making see *Behavioral Finance*, edited by Shefrin, 2001).

Research Paradigms and Methods

Behavioural economics and economic psychology test individuals' ability to determine what is best for them in choice settings such as lotteries, game settings, bargaining, and market settings. Usually, laboratory experiments are conducted. Davis and Holt (1993), Hey (1991), or Smith (1976) quite concretely prescribe how to conduct such experiments. For instance, participants in an experimental environment should behave as they do normally in the real world. To this end, the experimenter must establish an incentive structure, which in general proposes some type of reward medium to the subjects. Incentives should directly depend on the subject's actions, should not lead to satiation, and should be designed in such a way as to prevent all other factors that might disturb the subject's behaviour. Valid experiments also require participants to be honestly informed about the aims, and financial incentives are viewed as necessary to motivate participants to maximize their outcome. However, Bonetti (1998) finds little evidence to support the argument that deception must be forbidden, and Brandouy (2001) criticizes

the prescription of financial incentives, since they do not always appear sufficient to subjugate individual differences between participants.

Apart from the postulated prerequisites for 'good' experiments, suggestions are made for the measurement of risk and of the value of goods. Critical statements on the issue of measurement of risk have been made by Krahnen, Rieck, and Theissen (1997) and by Unser (2000; see also El-Sehity, Haumer, Helmenstein, Kirchler, & Maciejovsky, forthcoming). The economic value of goods, usually public goods, is usually measured in 'willingness to pay experiments' Given the increased use of this approach, its validity needs to be assessed. Ryan and San Miguel (2000) made simple tests of consistency in willingness to pay and found that approximately one-third of subjects failed the test to be willing to pay more for a good A than for another good B, given that they preferred A to B. Chilton and Hutchinson (2000), Morrison (2000), and Svedsäter (2000) also question the validity of the technique.

Further problems are seen in the experimental approach. The method in economic studies is to investigate people's ability to collect and adequately evaluate relevant information, in order to select the best alternative. However, even if people were able to collect and assess the relevant information, the question remains as to what is relevant. Sacco (1996) argues that individuals behave rationally if they are able to gather all the information relevant for the optimal solution of their decision problems and if they are able to process this information optimally. In spite of its apparent simplicity, this definition becomes problematic once the meaning of 'relevant' is questioned. In general, being based on meta-empirical assumptions, individual judgements of relevance are not *per se* evidence of the decision-maker's degree of rationality. While the efficiency of the information processing procedure is a relatively unambiguous notion in the theory of rational decision-making, the same cannot be said for the judgements of relevance of the information processed. Individual decision-making processes are necessarily based on a set of meta-empirical assumptions that Sacco (1996) calls 'subjective metaphysics'. Every decision-maker's beliefs are conditioned by an idiosyncratic, subjective metaphysics that informs the individual's causal psychology and consequently judgements of relevance. Lundberg (2000) argues that market agents, such as traders, analysts, commentators, etc. must make sense of the conditions before taking any potential action. The complexity and pace of markets make multiple explanations, often of diametrically opposite nature, highly likely. On the aggregate level, divergent views are held by market 'bulls' and 'bears', respectively. On an individual level, it is likely that each agent may maintain more than one explanation of the present, as well as more than one projection concerning future market developments. Lundberg (2000) therefore argues that understanding the processes of sense-making would be an important step.

Economic theory is outcome oriented with the assumption that profit is maximized. One prominent and articulate advocate arguing that outcomes

are not all that matters for economic welfare is Sen (1993). Anand (2001) too argues for the relevance of the underlying procedure, especially in interactions between economic agents. Callahan and Elliott (1996) criticize the main research focus on outcomes rather than on sense-making and exploration of conceptual systems to learn more about the way in which actors both share meaning and understand specific events and situations. Other researchers complain that there is too little qualitative research being done, such as focus groups (Chilton & Hutchinson, 1999), aimed at understanding people's motives and considerations. Sonnemans (2000) stresses the importance of studying how subjects reason in economic settings, and Callahan and Elliott (1996) argue that, despite gaining increasing legitimacy within the economic profession, experimental economics is mainly restricted to theory testing.

Anomalies in Financial Decisions

In behavioural finance and decision making, much research has been done on the task of theorizing how to pursue optimal strategies, incorporating scientifically based maxims (Shefrin, 2001) as well as advice from successful experts in the field. If investors behave rationally, as the hypothesis of efficient markets presupposes (see Fama, 1998), there would be no need for a behavioural theory. All investors would decide on Bayes' probability theory, guided by probability calculations based on all available information. Efficient market theory and its components involve perfect competition in financial markets. This means that investors behave as if they had no market power over prices. Markets are frictionless, implying that there are no transaction costs, taxes, or restrictions on security trading. In addition, all assets and securities are infinitely divisible and marketable. It is assumed that all investors have homogeneous prior beliefs and Bayesian expectations. All investors simultaneously receive the same relevant information that determines market prices. Moreover, individual preferences are restricted in the sense that all investors care only about the risk and expected return trade-off, and all investors are rational-expectations utility maximizers. The theory of efficient markets deals with perfectly rational traders who do not neglect the information relevant to price development. All such information is public, and there is no private information that can repeatedly be used in order to earn excess returns. Prices are assumed to reflect future profits, discounted to today's value. In practice, however, there are 'noise traders', that is, traders who behave irrationally, and markets exhibit imperfections, such as Monday and January effects, weekend effects, the emergence of bubbles, and crashes (Wärneryd, 2001).

Many attempts to explain inefficiencies in financial markets invoke irrational behaviour of market participants. While the prevalent assumption among financial economists is that irrationality is unpredictable, researchers

in behavioural finance have embraced ideas from cognitive psychology, according to which some deviations from rational behaviour can be explained and even predicted (e.g., Shefrin, 2001; Thaler, 1999). First, many studies on choice and decision making in practice suggest that genuine decision making is extremely rare. Most actions are routine or determined by habit (Katona, 1975). If choices and decisions are taken, non-rational people or 'noise traders' are frequently led by the behaviour of others, by their emotions and motives, they are overoptimistic and overconfident and perceive the situation as under their control, and they commit cognitive errors (Kirchler & Maciejovsky, 2002). Wärneryd (2001) summarizes a number of errors and biases that frequently occur.

Assuming asset traders are able to collect the available information and recognize what is relevant, are they then performing better than others who are not in possession of that information? Güth, Krahnen, and Rieck (1997) investigated to what extent insiders were able to exploit their advantage of being informed about an asset's fundamental value in a double-sided sealed-bid auction trading in high-risk assets. It was found that insiders could only partially make use of their advantage in terms of final portfolio values. Gerke, Arneth, and Syha (2000) also found no systematic overperformance of order book insiders. The authors conducted an experiment on the impact of order book privilege on traders' behaviour and the market process. A market participant who had insight into the order book received information about current trading opportunities and other traders' preferences and was thus privileged. Nevertheless, volatility and liquidity of the markets were not influenced heavily.

A particular problem for decision-makers is that they find it hard to follow the statistically prescribed probability updating, as postulated by Bayes' theorem. Updating can be defined as the process of incorporating information for the determination of probabilities or probability distributions. Incorporating new information presupposes the existence of some starting point: this is called prior knowledge. Updating can relate to probabilities or to parameters of distribution functions; that is, information can be used either to make estimations as to whether or not certain events will occur, or to make inferences about the parameters describing the process that generates such events. In the latter case, the estimated distribution can be used to calculate the probability that a certain event will occur. Updating involves the determination of new probabilities, given some new information. Behavioural scientists found that Bayes' rule is not necessarily efficient as a descriptive and predictive device. It is often misapplied or neglected in the process of updating. Huck and Oechssler (2000) found that subjects have difficulty in applying Bayes' rule correctly even if they are familiar with it, and appear to apply it by accident even if they do use it. Ouwersloot, Nijkamp, and Rietveld (1998) developed a model for measurement of the error in probability updating. In a laboratory experiment, the presumed

systematic impact of four characteristics associated with the messages given to the respondents was studied. For a large number of cases, probability updating resulted in outcomes that deviated significantly from Bayes' rule, and deviations were influenced by message characteristics such as precision, reliability, relevance, and timeliness.

Complex situations demand repetitive choices in a series of interdependent decisions where the state of the world may change, both of itself and as a consequence of previous actions. Wärneryd (2001) argues that decision-makers in such complex situations fail in goal formulation processes and are characterized by 'thematic vagabonding'. This is a tendency to shift goals: while trying to reach one goal, they shift to other goals, finally arriving at one that was never intended. In complex decision settings, when repetitive decisions are taken, dynamic problems solved, and intertemporal decisions made, it can hardly be expected that people will behave according to the theoretical solution of the problem at hand. Decision-makers are expected to work backward through the decision trajectories taking into account the principle of optimality. Anderhub, Güth, Müller, and Strobel (2000) and Müller (2001) show that, rather than applying backward induction, people reduce the complexity of the problem at stake and use heuristics to come to a solution. Investors make use of heuristics, such as representativeness, availability, anchoring and adjustment heuristics, that permit decision processes to be 'abbreviated', but can lead to biased estimations and judgements.

Additional anomalies result from the refusal to learn from experience, passivity, or external attribution of one's own failures. In hindsight, irrational decisions may be 'repaired' by reinterpretations and sense-making.

Prospect theory, endowment effect, and sunk costs

Since actors in financial markets need to form expectations about future developments, a central issue in decision-making is how individuals deal with alternatives that have insecure consequences. Prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) attempts to reconcile theory and behavioural reality in decision-making. It pays attention to gains and losses rather than to total wealth, assumes that subjective decision weights replace probabilities, and that loss aversion rather than risk aversion is an overriding concept. The human problem-solving process is assumed to involve two phases: editing and evaluation. The major component of the editing phase is assumed to be coding. This refers to the perception of outcomes as gains or losses relative to a subjective reference point, instead of in terms of the final state of wealth. Thus, contrary to the assumptions of standard economic decision theory, preferences are not invariant to different representations of the same problem. A further component of the editing phase is combination (i.e., simplifying choice options by combining the probabilities of identical outcomes). Segregation is the separation of the

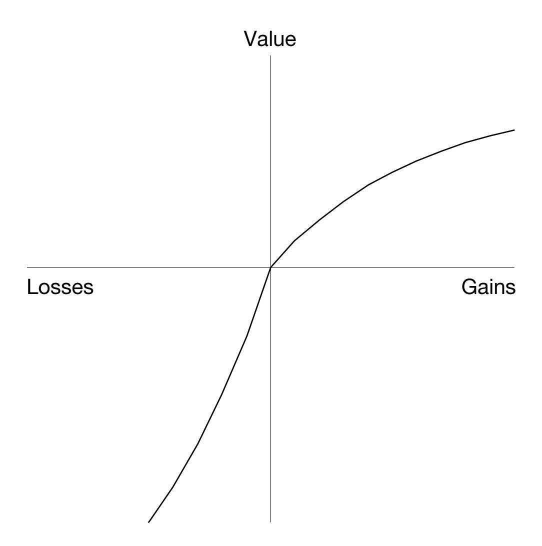


Figure 2.1 The value function (Tversky & Kahneman, 1981, p. 454).

certain component of lotteries from risky components. For example, the chance of winning a sum of 500 money units with probability $p = 0.80$ or of winning 300 units with probability $p = 0.20$ is decomposed into a sure gain of 300 and an 80% chance of winning an additional 200 units. Further components of the editing phase are cancellation, simplification, and detection of dominance. Probabilities of outcomes are often rounded to 'prominent' figures, and alternatives that are perceived as dominated by other alternatives are often rejected without further consideration. In the evaluation phase, decision-makers evaluate edited prospects, choosing the one with the highest value. Outcomes are defined and evaluated relative to a subjective reference point that represents the status quo of the individual's current wealth and marks the borderline between loss and gain. According to prospect theory, the value function is concave in gains and convex in losses, with additional gains or losses having diminishing impact. Furthermore, the function is steeper in losses than in gains, which indicates that losses loom larger than gains (see Figure 2.1).

The central implication of the value function is the basis for the principle of loss aversion. If losses are experienced or expected, people take risks to repair or prevent the loss; on the other hand, in gain situations decision-makers are risk-averse. Depending on the wording of a decision task, people perceive prospects as losses or gains and preference orders may be reversed. Such framing effects have frequently been confirmed. Departing from Tversky and Kahneman's (1981) Asian disease problem, Druckman (2001) presented participants with a survival format, a mortality format, or both. While the survival and mortality formats each replicated the original

experiment, Druckman (2001) shows how using both formats simultaneously provides a way to evaluate preferences unaffected by a particular frame and to measure the impact of frames relative to the baseline provided by the 'both formats' condition. For a critique on presentation formats of the Asian disease problem see, for instance, Kühberger (1995), Li (1998), and Wang (1996).

People buy insurances, but they also gamble and take investment risks. While people spend fortunes on lotteries, bets, and derivatives, the general assumption in economics is that people are risk-averse. Observations of how people deal with risks in real life have cast some doubt on the prevalence of risk aversion. People pay more than the expected value for insurance, and do likewise for lottery tickets. According to prospect theory, people are risk-averse for gains with high probabilities and for losses with low probabilities, risk-seeking for gains with low probabilities and losses with high probabilities. The tendency for people to be risk-seeking for gains with low probabilities is presumably enhanced as the sum involved increases. People will accept low probabilities to win large prizes in lotteries. The existence of large prizes with low probabilities would be more attractive than the possibility of winning smaller prizes with much higher probabilities. The most successful lotteries will then be those with high prizes and such low probabilities that the cost to the gambler can be held very low. Wärneryd (1996) investigated risk-taking in investments, playing in lotteries, and saving indices, and found that people who wanted to play safe asked for more than the required expected value. Most respondents showed risk aversion by preferring a prize that was certain to one that was probable.

Economic theory assumes that preferences are not affected by ownership. Thus, when income effects and transition costs are minimal, the amount a person is willing to pay for a certain good should equal the amount this person is willing to accept to give up this good. However, empirical research shows considerable differences between buying and selling prices of goods. Thaler (1992, p. 63) gives an example of what he calls the endowment effect: 'A wine-loving economist you know purchased some nice Bordeaux wines years ago at low prices. The wines have greatly appreciated in value, so that a bottle that cost less than $10 when purchased would now fetch $200 at auction. This economist now drinks some of this wine occasionally, but would neither be willing to sell the wine at the auction price nor buy an additional bottle at that price.' This apparently anomalous behaviour has stimulated much research, with most findings supporting the endowment effect. Stroeker and Antonides (1997) found that endowment effects can cause high reservation prices among sellers. Van Dijk and van Knippenberg (1996) conducted a market experiment with participants endowed with a bargaining chip that was convertible into real money after the experiment. The price was either fixed or uncertain, varying within a known range. Participants were allowed to trade their chips by offering and buying them

among themselves. It was found that selling and buying prices differed only in an uncertain situation, supporting an endowment effect arising from loss aversion. Selling prices significantly exceeded buying prices. In past studies on the endowment effect, people traded goods that were difficult to compare (e.g., coffee mugs, pens). Van Dijk and van Knippenberg (1998) studied the comparability of trading deals by exploring the relationship between the comparability of the gains and losses of the deal and the willingness to trade consumer goods. It was assumed and confirmed that people are more willing to trade wines from the same country than wines from different countries. People become more loss-averse with increasing incomparability of the gains and losses involved. Hoorens, Remmers, and van de Riet (1999) studied the endowment effect in the evaluation of time. Subjects indicated a higher figure for the payment they should receive for doing household and academic chores for others, than they considered fair to pay to the same others for doing identical chores. Disentangling the target and transaction dimensions that are usually intertwined in demonstrations of the endowment effect, two elements were found: subjects indicated higher fair wages for themselves than for another person (target effect) and higher fair wages for selling time than for buying time (transaction effect). The conclusion is drawn that the endowment effect in the evaluation of time rests on a combination of mere ownership (causing the target person effect) and loss aversion (causing the transaction effect). Mackenzie (1997) criticizes research on the endowment effect as concentrating mainly on behaviour, rather than the underlying motives. He argues that alternative motives may account for not trading wine in Thaler's example. Information about how people behave offers a low standard of evidence about what motivated their behaviour. In essence, a focus on motives is needed in place of drawing simple conclusions from behaviour to motives.

People tend to let their decisions be influenced by costs incurred at an earlier time. They are more risk-seeking than they would be had they not incurred these costs. This finding, the 'sunk cost effect', seems to be in conflict with economic theory, which implies that only marginal costs and benefits should affect decisions. Zeelenberg and van Dijk (1997) investigated the effect of time and effort investments (behavioural sunk costs) on risky decision-making in gain and loss situations. Participants were given vignettes to read, asking them, for example, to imagine that they had carried out a dirty job. They were then offered either payment of 50 money units or 100 units with a probability of $p = 0.50$, the alternative being 0 units (with the same probability). On the basis that gain or loss was to be fairly determined, participants were asked to state whether they preferred the certain payment or the gamble. A large number opted for the 50 money units, on the grounds that the expected frustration, if they received no reward after having carried out the work, was too great. However, if the participants were offered either 50 money units in addition to their pay or, in addition to their

pay, the chance to play a game that would bring them either 100 units or zero, each with a probability of $p = 0.50$, they chose the risky alternative. Realizing one alternative implies the loss of others, along with their consequences. In addition to risk propensity, anticipated regret (Loomes & Sugden, 1982) is a relevant factor in decision-making.

Capital asset pricing and portfolios

The theoretical economic models of portfolio theory and the capital asset pricing model offer a prescription of optimal investment behaviour. The idea is that, by integrating different shares into a portfolio, overall risk can be reduced below linear combination of single risks, unless assets are perfectly positively correlated. Diversification or widespread variance is the core of the model. Using information on the expected returns and expected risk of assets, the optimal mixture of assets resulting in an optimal portfolio for investors can be derived. However, as research in economics and psychology evidences, individuals and groups do not typically employ economically optimal behaviour. Various factors, such as limited information-processing capacity, lack of time, etc., are likely to impede rational decision-making.

A particular problem in portfolio allocation is the perception of risk. Measures of individual risk attitudes are required in financial management as well as in financial research. In asset management, for instance, investment allocation decisions will depend on clients' risk attitudes. Similarly, the analysis of individual portfolio choice in empirical and experimental research has to rely on an explanatory variable representing individual risk aversion. Various measures of evaluating risk aversion are in use. Krahnen et al. (1997) analysed certainty equivalents (i.e., the safe amount of money that leaves a decision-maker indifferent to a given lottery). In an auction experiment, the qualification of certainty equivalents as useful indicators of individual risk aversion was tested. Considerable deviation of evaluations over time was found. Krahnen et al. (1997) suggest using a multi-stage procedure, where parameter estimates are derived from a number of independent observations. Unser (2000) examined people's risk perception in a financial context and found that asymmetrical measurements of risk are superior to symmetrical risk measures such as variance.

Anderson and Settle (1996) investigated investment choice and the influence of portfolio characteristics and investment period. Portfolio allocation decisions require both fact judgements and value judgements. Fact judgements normatively require thinking about portfolio risk and return over a particular period on the basis of information about the mean, variance, and covariance of the component investments. Value judgements normatively require thinking about the consequences for lifestyle of various levels of return. The study focused on fact judgements and looked at estimates of return distributions at the end of an investment period, based on annual

return distributions. This issue is important since people have no real appreciation of exponential growth, having difficulty with non-linear functions. On the other hand, financial information is usually about annual measures of risk and return, investors usually invest for periods exceeding one year, and financial returns compound exponentially. Anderson and Settle (1996) further investigated estimates of return distributions of portfolios, based on return distributions of their constituents. It was found that people are insensitive to distributional characteristics in creating portfolios on the basis of both one-year and ten-year data. The authors interpret the results in terms of representativeness and anchoring-and-adjustment heuristics and a mental accounts decision rule.

Co-operation and Reciprocity

Research on choice in social settings shows the importance of justice and fairness considerations, and thus demonstrates that not only the consequences but also the process of decision-making is important. Neoclassical economic theory views altruism, co-operation, and reciprocity as non-rational in most circumstances, whereas self-interest and exploitation of others should lead to the highest profits. However, much research has provided empirical support for the robustness of social norms.

Recently, there has been a marked increase in literature claiming that there is more to economics than simple optimality; as Etzioni (1988) puts it, 'economics has a moral dimension.' This means that economic decisions take into account what is 'right' as well as what is most profitable. An interest in the role of morality or ethics in economic behaviour can be seen in studies of the importance of fairness and reciprocity (Fehr & Gächter, 1998; Fehr & Schmidt, 1999), of ethical values (Burlando, 2000), of business ethics (Wärneryd & Westlund, 1992), and in popular models of ethical and financial markets (Winnett & Lewis, 2000). Studies of ethical investment have been concerned with whether 'ethical' is mainly a marketing label, whether the performance of ethical trusts has a specific ethical component, and whether ethical investors are different from others and prepared to incur some costs in order to invest ethically. Webley, Lewis and Mackenzie (2001) show that ethical investors are prepared to choose ethical funds as part of a mixed portfolio as long as they are performing reasonably, but enthusiasm for investing ethically drops if the financial return is poor.

The norm of reciprocity and considerations of fairness are strong determinants of economic behaviour. Van der Heijden, Nelissen, Potters, and Verbon (1998) examined the force of reciprocity in gift exchange experiments in which mutual gift-giving was efficient but gifts were individually costly. The reciprocity norm had a powerful effect on behaviour. De Ruyter and Wetzels (2000) report the strong effect of pro-social behaviour in the marketing context with soccer fans buying shares from their club in order to provide

assistance in times of financial need. Church and Zhang (1999) found in a bargaining study that the majority of subjects were concerned about maximizing their pay-offs, but the second most frequent response was being fair to the other. Gneezy, Güth, and Verboven (2000) found that people in long-term contractual relationships, which can never be specified in all details, make decisions that are based on trust and reciprocity.

The basic assumptions and propositions of classical economic theory are frequently investigated using ultimatum games (Güth, Schmittberger, & Schwarzer, 1982). In the simplest version, one player (the allocator) is directed to divide a sum of money (the stake) between himself and a second player (the recipient). If the recipient accepts the proposed division of the stake, then both receive the amounts proposed by the allocator. If the recipient rejects the proposed division, then both players receive nothing. The recipient knows the size of the stake, but the players do not know each other. According to the model of Homo oeconomicus, allocators should offer the smallest possible amount that makes the recipient better off than nothing. For instance, if 100 money units are at stake and divisible into units of 1, then offering 1 unit makes the recipient better off than nothing and he should accept the division. The allocator would get 99 units, whereas the recipient gets 1 unit.

Several empirical studies strongly confirm that allocators are not as selfish as economic theory predicts, and recipients are not willing to accept the lowest possible amount. Huck (1999) investigated responder behaviour and motives such as equality, self-esteem, consideration of absolute and relative pay-offs, malevolence, fairness, and revenge. Five groups of responders with different motivations guiding their behaviour were detected: (a) malevolent subjects purely guided by their absolute pay-off; (b) malevolent and highly competitive subjects, who care for their own fair share in ultimatum bargaining and are willing to sacrifice a substantial amount of money to increase their relative pay-off; (c) non-malevolent subjects whose only concern is for their absolute pay-off. They prefer even small amounts of money to receiving nothing by inducing a conflict. (d) Non-malevolent but vain subjects who differ from cluster (c) only by rejecting 'peanuts'; and (e) non-malevolent participants with a real desire for equality, for whom fairness considerations matter most. Bethwaite and Tompkinson (1996) focused on motives that drive players in ultimatum games to offer, accept, or reject certain amounts: fairness, envy, altruism, or selfishness. Their study found the dominant motive to be fairness. Half the recipients had a concern for fairness; only one-quarter were motivated by selfishness and so had a utility function of the type conventionally assumed by economists.

The modal offer in ultimatum games is usually not the lowest positive amount, but the even split of the pie. While one explanation of such behaviour invokes the notion of fairness, a second explanation is that in the absence of common knowledge of the rationality and beliefs of recipients, allocators

raise their offers because they expect that unsatisfactory offers might be rejected. In several experiments, selfish offers were successfully induced by manipulating the allocators' expectations. Suleiman (1996) argues that such manipulations are predominantly extrinsic and not accounted for by game theory. In his study, he introduced a minor variation of the ultimatum game by implanting a discounting factor in the standard game: if the recipient rejected an offer, the offered division was multiplied by a factor ranging from 0 to 1. The change was that, instead of receiving nothing at all if the recipient rejected the offer, the players at least received something by acceptance. Whereas game theory is indifferent to this modification, experimental results from the modified game showed that by continuously changing the discounting factor, it is possible to induce systematic changes in allocators' and recipients' behaviour and beliefs. The results show that the allocator is driven by expectations about the recipient's behaviour, but, at the same time, norms of fairness cannot be ruled out.

Experimental evidence indicates that individuals often exhibit other-regarding behaviour when bargaining with other people. Violation of the presumption of self-interest in neoclassic economic theory has promoted intense debate within behavioural economics, and many experiments have been conducted to detect conditions that push allocators to behave rationally (Cherry, 2001). Güth, Ockenfels and Wendel (1997) investigated co-operation based on trust in a basic sequential game, the so-called 'trust game'. The first mover starts by deciding between co-operation and non-co-operation, while the second mover can only react in the case of co-operation, either exploiting the other player or dividing the rewards equally. Trust and co-operation were shown to depend on how the positions of players in the game were chosen. Sonnegård (1996) tested whether the random choice of movers determines the amount given. While random choice had no effect, description of the property rights of the first mover determined giving. If the first mover was explicitly instructed to have the right to exploit the other player, then they offered less. In addition, high incentives led to less co-operative behaviour. If small stakes are to be divided and sequential games are played, individuals may try to explore the partner's acceptance level by varying the offered percentage of the stake (for exploration and learning in decision settings as well as boundedly rational decision-making see Güth, 2000; Albert, Güth, Kirchler, & Maciejovsky, 2002; Mitropoulos, 2001).

Self-interested behaviour may also increase when the stake to be divided is not known to the recipient. Murnighan and Saxon (1998) conducted ultimatum games with children and found that younger children make larger offers and accept smaller offers than older participants. Boys in the age group nine to fifteen years seem to take greater strategic advantage of asymmetric information than girls. Like adults, children accepted smaller offers when they did not know how much was being divided.

Why are people co-operative and take so much account of fairness norms?

Scharlemann, Eckel, Kacelnik, and Wilson (2001) argue that, although economists and biologists view co-operation as anomalous, since animals that pursue their own self-interest have superior survival odds to their altruistic or co-operative neighbours, in many situations there are substantial gains to the group if members co-operate. Individuals reap the benefits of co-operation if they are able to detect the intention in others to co-operate. For instance, smiling may be a signal of willingness to co-operate. Scharlemann et al. (2001) conducted a two-person, one-shot trust game with participants seeing a photograph of their partner either smiling or not smiling. Results lend some support to the prediction that smiles can be a signal of co-operation and can elicit co-operation among strangers. Trust was correlated with smiling but was more strongly predictive of behaviour than a smile alone.

ECONOMIC SOCIALIZATION AND LAY THEORIES

Children's Economic Knowledge and Behaviour

A knowledge and understanding of economic cause and effect assumes a process of maturation and socialization. Pre-school-age children know little of the production and distribution of goods, supply and demand, or other economic systems. Knowledge is still slight at the age of ten to eleven years, and a differentiated understanding cannot be assumed until the age of about 14 years.

Jean Piaget (1896–1980) developed a theory of the development of human intelligence that is also useful to describe the development of economic knowledge (Berti & Bombi, 1981; Kirchler, 1999). Piaget assumes that the development of intelligence is a process aimed at achieving and maintaining harmony of the individual and the environment. Knowledge can only be achieved by concerning oneself with an object. This concern with an object, whether concrete or imagined, brings about transformations, an act of adaptation. Adaptation is a fluid state of equilibrium between conforming or assimilating the environment to the person and conforming or accommodating the person to the environment. For example, there is an attempt to explain new and unfamiliar facts on the basis of the models or mental frameworks available to the individual. Assimilation processes involve the integration of unknown information into the available frameworks. Grappling with previously unknown facts eventually leads to a deeper understanding and to the differentiation and adaptation of the subjectively available explanatory models to the new facts, an act of accommodation. Cognitive development is a process of achieving increasing harmony between the assimilatory and accommodatory exchange processes between the individual and the environment. There is an associated process of generalization, differentiation, and co-ordination of the cognitive structures. The exchange processes between the individual and the environment enable individuals to proceed from an

initial global condition to a cognitive structure that is enduring, flexible, and organized in a differentiated way, and permits logical thought.

Comprehensive studies on the development of children's economic knowledge have been carried out in Italy by Berti and Bombi (1981). In this area, too, as in Piaget's theory on the development of intelligence, children apparently begin with merely diffuse, unrelated economic concepts. They have little knowledge of the production of goods. Work, income, consumption, etc. are viewed as separate concepts and not related to each other. Increasing age brings comprehension of increasingly complex relationships, but only when children reach about the age of 14 years do they begin to develop a clear, complete picture of basic economic matters. Similar findings have been reported from various other countries (see *Journal of Economic Psychology*, 1990, Issue 4). Developmental stages in line with Piaget's theoretical model have been demonstrated in Hong Kong, North Africa, Europe, Australia, and North America. Bonn, Earle, Lea, and Webley (1999) investigate children's views of wealth, poverty, inequality, and unemployment in South Africa. The results show that the capacity to make inferences and integrate information about these concepts is most influenced by age, but that the particulars of the children's knowledge are influenced by their social environment. The process of knowledge acquisition can be accelerated by experience and training. Children in deprived areas and children from poorer families achieve differentiated economic knowledge earlier than other children, on account of the need to work for a low wage or to deal with materialistic options at an earlier stage. Further, talking to children about economic matters (Cram & Ng, 1994), training and instruction improve economic knowledge on the part of children and young people (Aquino, Berti, & Consolati, 1996).

Children attain economic knowledge, and they are addressed as appropriate economic partners by advertisers. In some cases, children and young people are given a considerable say in joint purchasing decisions (see Kirchler, Rodler, Hölzl, & Meier, 2001). In some cases, they have autonomous control of sizeable sums of money. Children receive pocket money, presents of money, and modest financial reward for minor tasks in the home. According to Furnham (2001), over 88% of parents questioned in his study thought that children should receive pocket money from the age of about six years. The amount of pocket money increases linearly with age, and children spend and indeed save part of their money, independently of the funds at their disposal. On average, boys receive more pocket money and bigger presents than girls (Furnham, 1999, 2001).

Lay Theories

While the representations of the experts about their knowledge are detailed and logically structured, lay people tend to rely on everyday experience to

construct cognitive frameworks by which to plan and rationalize behaviour and integrate new information into their existing fund of knowledge. Although Furnham (1997) maintains that in investigating work and economic values, people have coherent socio-economic, ideological belief systems, this in fact merely signifies that the 'subjective theories' are consistent in themselves, but serious differences exist between individuals and it is difficult to generalize.

It must, however, be stressed that even economists' knowledge or the resulting political advice on economic matters need not be unified. The principle that individuals construe reality in different ways is also evident in the way that politicians and economic experts construe policies (Theodoulou, 1996) and their implications. Different individuals have different systems of thought; these may be viewed as different strategies, which they consistently use to make sense of the world. Theodoulou (1996) studied the way in which Labour, Conservative, and Liberal Democrat experts in economics and business matters construe economic and political reality. She found significant differences between Labour and Conservative Party supporters in their preference for propositional, aggressive, pre-emptive, and hostile construing.

Lay theories have been investigated mainly from the point of view of attribution theory and the theory of social representations. Following the fall of communism in Eastern Europe, Antonides, Farago, Ranyard, and Tyszka (1994) and Tyszka and Sokolowska (1992) investigated lay conceptions of economic values and desires. Williamson and Wearing (1996) interviewed 95 individuals about the present state of the economy and expectations over the short and long term, confidence in organizations involved in the Australian economy, current information about the economy, and other related issues. The authors detected as many unique cognitive models as individuals interviewed. Despite the differences between them, there were some broad areas of agreement. In general, individuals described the economy by integrating economic, social, psychological, and moral issues. In some respects, previous findings suggesting that lay people know little about fiscal issues were confirmed. However, the cognitive models showed that lay people did seem to understand some connections between government revenue and expenditure.

Specific topics include above all lay concepts of poverty and affluence and their subjective causes. Poverty and wealth are often attributed to internal causes. Christopher and Schlenker (2000) studied perceived material wealth. Participants in their study were given vignettes to read that described a person in either an affluent or not so affluent home setting. The affluent target was evaluated as having more personal ability, such as intelligence and self-discipline, was perceived as having more sophisticated qualities, for example, as being cultured and successful, and as having a more desirable lifestyle than the not so affluent target. However, affluent people were judged

as being less kind, likeable, and honest. Studies examining the individualistic, fatalistic, and structural dimensions of poverty showed that the majority of Americans explain poverty in individualistic terms (Hunt, 1996). Abouchedid and Nasser (2001) examined causal attributions of poverty among Lebanese students and found higher ratings for structural explanations than for individualistic ones.

Other specific topics relate to unemployment, borrowing and debt, the burden and equity of taxation, and personal consumption. The conclusions generally confirmed that lay people believe in a 'just world': respondents thought that people in difficult situations are usually themselves to blame (Kirchler, 1999). As regards the state, citizens in many countries are sceptical. Taxes in particular are not always seen as a necessary evil. Opinion surveys show that most people want a reduction in taxes, but, inconsistently, at the same time favour a rise in state spending in almost all areas. The public welcomes the benefits of public goods, but is becoming ever more unwilling to pay the cost (Tyszka, 1994; Williamson & Wearing, 1996). Kirchler (1998) investigated the attitudes of the self-employed, entrepreneurs, public employees, students, clerical and blue-collar workers to taxes. They found not only a general suspicion that taxes were not being used appropriately, but fear that the distribution of the burden was neither horizontally nor vertically fair. For entrepreneurs and the self-employed, who pay taxes 'out of pocket', there was also the problem of the experience of loss and the sense of demotivation and the restriction of freedom.

A glance at consumer behaviour shows that purchasing behaviour permits some interesting conclusions about subjective theories, values, and desires. First, subjective images of the economy direct people's actions, and, second, behaviour is a source of information about the person. Dittmar and Drury (2000) emphasize the role of personal consumption as providing a picture of the person: 'The self-image is in the bag!' Impulsive buying, consumption, and regret have complex meanings beyond those that can be measured easily in survey research. Janssen and Jager (2001) emphasize the need for identity, which explains in part the dynamics of consumer markets. Kasser and Grow Kasser (2001) investigated the dreams of people with high and low levels of materialism. People high in materialism reported more insecurity themes (e.g., falling), more self-esteem concerns, and dreams about conflictual interpersonal relationships. People with low materialism reported dreams where they were able to overcome danger, and typically moved toward greater intimacy in their dreams. The authors interpret their findings as supporting the notion that feelings of insecurity might be connected with the pursuit of material values. Consumption as economic behaviour is an expression of individual values and desires, and subjective constructions of the self, society, and the economy.

ENTREPRENEURSHIP AND ISSUES OF ECONOMIC PSYCHOLOGY IN THE COMMERCIAL COMPANY CONTEXT

Entrepreneurs, commonly seen as sensitive to economic and social phenomena, far-sighted, and prepared to take risks, fundamentally shape and innovate economic life by their activities. From the perspective of economics, the role of the entrepreneur is strictly determined by market transactions, and entrepreneurs have no freedom to develop their own individual characteristics. There would therefore be little relevance in devoting scientific attention to the personality of the entrepreneur. From a psychological perspective, the particular tasks and activities of the entrepreneur do, however, imply that certain personality traits are relevant for initiative and exercise of the enterprise function, and that it should be possible to find psychological distinctions between entrepreneurs and other market participants. In particular, personality and motivational structures, religious convictions, and value concepts have been investigated as supposed determinants of entrepreneurial success (Kirchler, 1999).

If entrepreneurs are to make independent decisions, take action, and bring innovative ideas into effect, they should carefully analyse business-relevant information. Bailey (1997) found that managers with greater ‘need for cognition’ produced a more thorough information search in the judgements of job candidates. It can further be presumed that entrepreneurs need to be independent of others, prepared to take risks without being blind to risk, interested in social contacts, and emotionally stable. Brandstätter (1997) studied owners of small and medium enterprises and people interested in setting up a private business. A personality checklist showed that owners who had personally set up their business were emotionally more stable and more independent than those who had taken over their business from parents, relatives, or by marriage. People interested in setting up their own business had similar personality characteristics to founders. Individuals who had personally founded their businesses or planned to do so, and who achieved high scores for independence and stability, were happier with their role as entrepreneurs, more satisfied with past success, more confident of future success, more inclined to attribute success to internal causes, and more likely to be thinking of expanding their business than those who scored low for those factors. Korunka, Frank, and Becker (1993) also report high independence scores for successful business founders. It remains speculation whether the features in the personality questionnaires can indeed be causally linked to business success. The personal self-descriptions of the respondents may be partly reality, partly wish, partly the cause, and partly the effect of past business experiences.

Various findings exist as to entrepreneurs’ readiness to take risk. On the one hand, high risk propensity is seen as a precondition of innovation; on the

other, precipitate action can be commercially disastrous, and the careful weighing of the consequences economically wise. Only where the probability of success of a particular action is sufficiently high is it worth taking a certain risk. In a three-dimensional system with the dimensions of innovativeness, reasoned goal-directedness, and risk propensity, Wärneryd (1988) places the successful entrepreneur at that point where both the inclination to strike out on new paths and the desire for success are particularly high, and where the readiness to bear a certain degree of risk in the case of success-promising actions is also above average. Actions with little promise of success are avoided. Successful entrepreneurs appear to give particularly careful consideration to risk. Frank and Korunka (1996) stress that risk propensity must be analysed in a situation-specific context: in their study, successful entrepreneurs were particularly oriented toward action and decision in the face of possible failure, but persistent in holding on to the situation in success.

Entrepreneurs and managers should be able to make reasonably accurate economic prognoses. Anderson and Goldsmith (1997) investigated managers' profit expectations, the degree of confidence placed in their profit forecasts, and investment levels. In most industries, investment increased both when managers were more optimistic and when they exhibited greater confidence in a forecast. Aukutsionek and Belianin (2001) studied the quality of forecasts and business performance by Russian managers. Forecast quality was typically poor, and most managers exhibited overconfidence by judging their forecasts as good.

A particularly relevant decision to be made by entrepreneurs is whether to wait or to become active and found a company. Davidsson and Wiklund (1997) show that values, beliefs, and culture have an effect on regional new firm formation rates. In discussing the market entry decision and the interaction between an incumbent firm and a potential entrant, the focus in the literature has been on two aspects: the strategic implications of having a first-mover advantage, and the different asymmetries that may be created by the incumbent, for example, cost asymmetries, capacity asymmetries, brand loyalty, or any other factor that affects the firm's profit functions. Important managerial decisions such as entry into new markets and exit from existing markets are made by people, and people are characterized by bounded cognitive abilities. Facing seemingly similar decision problems, individuals might evaluate them differently and therefore might come to different conclusions. Fershtman (1996) analyses incumbency using prospect theory, and explains firm decisions by the company's reference points. The managers of the incumbent company evaluate decisions from the point of view of being within the industry, while, for the management of the entrant, the reference point is that of being outside the industry. The difference in the reference points leads to different market decisions.

Varying reference points and loss aversion are also relevant for changes in management. Replacing one manager by another, even with the same

qualifications, may have an important effect, as it introduces a manager with a different reference point. For example, following a recent loss, a manager might retain the reference point held prior to the loss, since any adjustment of reference points is not necessarily immediate. Replacing the manager may induce different managerial behaviour simply because the new manager may refer to the new status quo as his reference point. Another interesting point is the possible effect of sunk costs: a manager with a particular point of reference might attempt to repair losses by remaining in the market, whereas a newly appointed manager might take the status quo as a starting point, see no hope for the future, and leave the market. A further effect of loss aversion is inaction inertia. Participants who fail to act on an initial opportunity are less likely to act on a second somewhat less desirable opportunity compared with participants who did not experience inaction (Tykocinsky, Pittman, & Tuttle, 1995). Butler and Highhouse (2000) showed that decisions to sell a corporation are less likely after failing to act on a previous offer.

Entrepreneurs and managers often make decisions under time pressure and in the face of inadequate information about alternatives and consequences. Economic theories of the firm assume rational decisions; Wakely (1997) proposes a model using the theory of bounded rationality and shows that managers, starting from their aspiration levels, aim at satisfying results and not at optimal solutions. Kristensen and Gärling (1997) prove that, in commercial transactions, considerations of fairness, the prospect of further co-operation, and the build-up of trust are of greater significance than the rational model would imply. The variables of profit-orientation and fairness are also relevant in the interaction between consumers and commercial companies. A company that is solely profit oriented risks both image and customer loyalty. Seligman and Schwartz (1997) studied the role of fairness in economic situations by referring to Kahneman, Knetsch, and Thaler (1986). Respondents made fairness judgements with the aim of deriving descriptive generalizations of people's intuitions about the fairness of companies in economic transactions. Typical questions asked respondents to judge the fairness of an imaginary commercial company's action, as in the following example: 'A hardware store has been selling snow shovels for $15. The morning after a large snowstorm, the store raises the price to $20.' The results confirmed the findings of Kahneman et al. (1986) on fairness judgements made for companies. However, they also demonstrated that people judge parallel actions by individuals as fair. People apply different standards to individuals and companies because of presumed differences between them in wealth, power, and size. When companies are portrayed as no more powerful or wealthy than individuals, differences in fairness judgements were eliminated. Further, respondents were less inclined to judge the behaviour of a commercial company harshly when the company was identified with an individual than when it was large and anonymous.

Work Experiences and Income

Work experiences are mainly investigated by occupational and organizational psychology. Economic psychology concerns itself primarily with pay as a factor for motivation and productivity, and with the factors determining pay.

Tang (1996) investigated the acceptance and justification of differing levels of pay. When allocating money to different positions, men who value money highly have a strong preference to reward those who occupy the highest positions and to offer very little to those in the lowest ones, while no significant differences were found for women. For men with positive attitudes toward money, those who have power and authority deserve to have more money than those who do not.

What determines the level of income? Plug (2001) asked whether schooling pays off with regard to income. This seemingly simple question puzzles many economists and no full answer appears to be known. Groot and van den Brink (1999) investigated work stress from an economic perspective and the monetary equivalent of stress. Evidence was found that men report stress more frequently than women and there is a sizeable compensation for work with stress. Workers in jobs with stress earned 6–9% more than they would have earned in jobs without stress. Physical attractiveness was also studied as a determinant of income (see Bosman, Pfann, Biddle, & Hamermesh, 1997; Kyle & Mahler, 1996). Schwer and Daneshvary (2000) report that, in general, less attractive people earn less than better looking people. Attractiveness was found to influence income attainment for men, older individuals, and those employed in predominantly male occupations and in occupations that rely on person-to-person contact, and in which appearance may influence economic productivity. The starting salary of men was significantly influenced by their attractiveness, but not so for women. However, both attractive women and men earned more over time. Schwer and Daneshvary (2000) found that persons employed in occupations in which appearance could influence job performance frequented other types of hair-grooming establishment and attached more importance to their appearance.

From a managerial perspective, paying a fair wage is important because workers evaluate their compensation by comparing it with that of others of similar standing. Motivation to work, satisfaction, and organizational commitment depend on fairness perceptions. From a purely economic perspective, wages should be as low as workers can just accept. Fehr, Kirchler, Weichbold, and Gächter (1998) and Kirchler, Fehr, and Evans (1996) contrasted the implications of standard economic theory with social exchange predictions. According to standard economic theory, workers and employers are rational, egoistic individuals who strive to maximize profit. In markets, as well as in bilateral interactions, employers should offer the lowest wages that

workers will accept and workers should provide the effort level that maximizes their utility (i.e., the minimum permitted). According to social exchange principles, wage negotiations between employers and workers are not only determined by egoistic profit maximization but also by social norms, such as reciprocity. Employers are assumed to trust reciprocation norms and offer higher than reservation wages, expecting workers to provide higher effort in response. Consequently, workers' effort choices are expected to be positively correlated to employers' wage offers. Reciprocation norms were found to be important and, on average, co-operation was considerably higher than predicted by economic theory. However, there were significant differences between participants: some workers reciprocated higher offers over a series of bilateral trading periods and in market situations, whereas others did not. Besides social norms, altruism, reciprocity, and competition motives were important. Falk, Gächter and Kovács (1999) investigated opportunities for social exchange in games with incomplete contracts and found similar evidence for the importance of fairness and reciprocity.

Workers work hard for money and harder for more money. Goldsmith, Veum, and Darity (2000) studied the efficiency wage hypothesis, which states that firms are able to improve worker productivity by means of a wage premium (i.e., paying a wage above that offered by other firms for comparable labour). A link between wage premiums and productivity might arise for a number of distinct reasons: a wage premium may enhance productivity by improving nutrition, boosting morale, and encouraging greater commitment to company goals; it may reduce leaving rates and the disruption caused by turnover, attract higher quality workers, and inspire workers to greater effort. Goldsmith et al. (2000) used locus of control as an index of effort and found support for the efficiency wage hypothesis.

To a large extent, people judge their personal welfare by comparing it with others within their local environment. Workers' tendency to evaluate their compensation by comparison with other, similar workers has been a topic of much empirical and theoretical interest. Not only money and effort, but also non-monetary compensation such as work status is compared. Schaubroeck (1996) argues that an understanding of organizational attachment may be facilitated by examining the local hierarchies in which workers trade off income for status. Within this perspective, individuals seek alternative employment when the income–status balance is not to their liking. Frank (1985) argued that the utility of a given pay level is a function of its rank within the organization (i.e., 'pay status') and the absolute level of the pay. Higher pay confers status relative to others in the organization because it signals the importance of individuals to the organization as well as their ability to acquire positional goods. Individuals will therefore choose their employing firm based on their relative preference for status.

Unemployment

Psychologists have offered theories to explain how experiences of joblessness may lead to a decline in mental health in general and various aspects of emotional health and self-esteem in particular. Goldsmith, Veum, and Darity (1996) claim, however, that omitted variables, unobserved heterogeneity, and data selection have prohibited the emergence of a consensus on the impact of unemployment on self-esteem. They investigated data from the US National Longitudinal Survey of Youth that provide detailed information on the personal characteristics of individuals in the sample, including self-esteem as well as their labour force experience. Goldsmith et al. (1996) found clear evidence that joblessness damages self-worth. Unemployment significantly harms self-esteem, and the effect of such exposure persists.

Unemployment is a critical life event leading to lack of self-confidence, helplessness, hopelessness, inefficiency, fatalism, fear of the future, and depressed mood in both Western industrialized countries and developing countries. The unemployed, educated young men in India who took part in a study conducted by Singh, Singh, and Rani (1996) on self-concepts of unemployed had generally rated themselves relatively low, though moderate, on variables concerning private and social self. In addition, most participants indicated suffering a considerable amount of social conflict.

While the negative impact of unemployment on psychological health is well known, less is known about how people cope with the problems associated with unemployment, one of which is economic deprivation. Waters and Moore (2001) examined the interrelationships between employment status, economic deprivation, efforts to cope, and psychological health. The results suggest that economic deprivation is experienced differentially in respect of material necessities and meaningful leisure activities, with unemployed respondents differing from employed on levels of deprivation for meaningful leisure activities but not for material necessities.

A topic of interest in labour economics is the extent to which past unemployment has an effect on current labour market status. Empirical results on unemployment hysteresis are, however, contradictory. Darity and Goldsmith (1993) utilized social psychological research on the effects of unemployment to explain unemployment hysteresis. Unemployment gives rise to a general sense of helplessness, and it is reasonable to conclude that this sense of not being in control will also be positively correlated with the length and frequency of spells of unemployment. Elmslie and Sedo (1996) developed an economic model of unemployment hysteresis based on social psychological findings, especially learned helplessness theory. They showed that perceived labour discrimination leads to several adverse psychological conditions that impair an individual's human capital characteristics such as learning abilities and motivation, resulting in turn in decreased future employability. In this way, unemployment ultimately has high social economic costs.

HOUSEHOLD FINANCIAL BEHAVIOUR

Purchase decisions as spontaneous, habitual, extensive individual decisions, or joint decisions between partners and their children are a relevant research field in marketing and consumer psychology with a long tradition (Kirchler et al., 2001). In addition to partners' spending behaviour, economic psychology concerns itself with their savings decisions (e.g., Wärneryd, 1999), decisions relating to credits and debts (e.g., Walker, 1996), loan and loan duration estimations (Overton & MacFadyen, 1998), including loan credibility judgements (Rodgers, 1999), money management in general, and open-mindedness toward innovations in the money sector, like wall-banking (Pepermans, Verleye, & van Cappellen, 1996).

Traditional microeconomic theory focuses primarily on the behaviour of the individual, whereas microeconomic applications focus on the behaviour of the household. The maintained approach is to assume that the household acts as an individual and has a unique welfare function. Katona (1975) established the importance of social perception variables in mediating the impact of economic conditions on financial decision-making. Lay perceptions of the economy are important mediators between economic conditions and economic behaviours. People respond to their perceptions of present and future economic conditions rather than directly to objective features of the economy, and these perceptions are also heterogeneous between individuals when living in the same household. However, Plug and van Praag (1998) found considerable similarity in the household members' construction of family equivalence scales of welfare. Where the age and educational characteristics of partners in the household are identical, then, according to Plug and van Praag (1998), a single welfare approach can be justified for the description of household response behaviour. From a social psychological and consumer psychological perspective (Kirchler et al., 2001), it is, however, problematic at the very least to assume that husband and wife and their offspring have similar views of the dynamics of spending and saving behaviour. Approximately one-third of responses of partners differ when they are asked who influences what decisions, and how decisions are reached. Viaud and Roland-Lévy (2000) apply social representation theory to study consumption in households when facing credit and debt, and find different intra- and inter-household constructions of money matters.

The interest of economic psychologists in household savings behaviour has increased in recent years. Wärneryd (1999) has produced a comprehensive survey of the literature on saving, and a special issue of the *Journal of Economic Psychology*, 1996, is dedicated to household savings behaviour. While, in economics, saving is mainly analysed within the framework of the life cycle hypothesis, Wärneryd (1999) emphasizes the importance of psychological phenomena, such as attitudes, expectations, and subjective concepts of savings in general.

Analysing savings discourses, Lunt (1996) found that people base their understanding of economic change on broader conceptions of psychological, social, and political change. People are aware of the increased opportunities made available through the deregulation of the financial markets and the shift in institutional forms of banks. They are also aware that this can lead to chances and dangers for the consumer. Changes led to a new climate for consumption marked by increased individual responsibility for insurance (as an investment rather than for risk reduction: Connor, 1996), as opposed to institutional methods, along with increased uncertainty over both the methods of insurance and the present and future risks that the individual and family face.

Households are extremely risk-averse, but still the degree of risk aversion varies considerably between households. Pålsson (1996) examined household risk-taking in Sweden. Based on a standard model of intertemporal choice, relative risk aversion can be expressed in terms of the proportion of total wealth invested in high-risk assets and the price of risk. Since households tend to construct different portfolios, consideration was taken not only of differences in the proportion invested in high-risk assets, but also of differences in the composition of the high-risk assets. Pålsson (1996) showed that aggregate, relative risk aversion coefficients are generally high and increase with the age of household members. Donkers and van Soest (1999) analysed three subjective measures of household preferences that can influence the household's financial decisions: time preference, risk aversion, and interest in financial matters. The relations between these variables, family income and family characteristics, and financial behaviour related to housing and ownership of high-risk financial assets were assessed. Risk aversion was negatively correlated with decisions to invest in high-risk financial assets. Also, risk aversion increased with age, and women were more risk-averse than men. Households with higher incomes and men were found to be more interested in financial matters, and consequently more likely to own high-risk assets.

What forms of saving do households choose and what strategies do they use? Groenland, Kuylen, and Bloem (1996) found savings related to banking options (e.g., savings accounts), old age (e.g., pension schemes), durables and own property (e.g., buying a house). Three characteristics of saving seem to be relevant for savings decisions: saving may be contractual or non-contractual, interest on savings may be fixed or non-fixed, and saving may be aimed at increasing one's personal wealth or at maintaining the value of one's capital over time. Wahlund and Gunnarsson (1996) studied phenomena of mental discounting across households and found specific savings strategies. In two additional studies, the authors identified residual savers, contractual savers, security savers, risk hedgers, prudent investors, and divergent strategies. Residual saving strategies were found to be the most frequent, followed by contractual saving, security saving, and risk hedging. The observed variation in preferred savings strategies depended on time preference measures, degree

of financial planning and control, interest in financial matters, attitudes toward financial risk-taking, propensity to save, and financial wealth (Gunnarsson & Wahlund, 1995, 1997).

The motives for saving differ across cultures (Jain & Joy, 1997) and households, and seem to vary depending on personality characteristics (Brandstätter & Güth, 2000). It can be assumed with regard to influence in joint decisions about savings and money matters that the more a partner is interested and expert on an issue, the higher this partner's influence in joint decisions (Kirchler et al., 2001). Meier, Kirchler, and Hubert (1999) report that in savings decisions and decisions about investments in assets, the expert partner is more influential, and men are generally more influential in partnerships with traditional role orientation.

MONEY AND THE EURO

An exact definition of money is hard to give, and Snelders, Hussein, Lea, and Webley (1992) term money a 'polymorphous' concept. Rumiati and Lotto (1996) asked experts, such as bank clerks, and non-experts to judge the typicality of a list of money exemplars. Three factors emerged: ready money (e.g., coins and banknotes), bank money (e.g., cheques, bank drafts, credit cards, bank cards) and money substitutes (e.g., vouchers, telephone cards), with the first factor prototypical for money. From an economic psychological perspective, in recent years the use of credit cards (Hayhoe, Leach, & Turner, 1999), acceptance of wall-banking (Pepermans et al., 1996), and attitudes toward money (Lim & Teo, 1997) were in the focus of research.

Another issue is the subjective value of money and the subjective perception of prices. With regard to price perception, Kemp and Willetts (1996) found that people who estimated present, past, and future prices of wool, butter, stamps, and general living costs, have no correct memories of prices. More recent prices are frequently underestimated, while overestimations are likely to occur when subjects are asked to estimate prices more than a decade ago. Brandstätter and Brandstätter (1996), asking what money is worth, report that the utility of money is a function of income (low income, high utility) and personality characteristics, such as extroversion and emotional stability. When subjects were asked to categorize levels of annual income as poor, nearly poor, etc. to prosperous, income and family size were important determinants. Using the method of just noticeable pay increments, Hinrichs (1969) asked employees what amount of pay increase they would rate on a five-point scale ranging from 'barely noticeable' to 'extremely large'. The marginal utility of the same additional amount of money should be higher for low-income individuals than for rich people. According to Weber's law, a just noticeable difference, measured in physical units, is a constant

proportion of the magnitude of the standard stimulus, expressed in the same physical units. This was roughly true in Hinrichs' survey. However, Champlin and Kopelman (1991) and Rambo and Pinto (1989) failed to replicate these findings. Brandstätter and Brandstätter's (1996) study tested the validity of Steven's power function and the influence of monthly net income, attitudes toward money, and personality traits on the subjective value of money. People had to imagine winning or losing certain amounts of money and to indicate their resulting emotions, such as joy and anger. It was found that the subjective value of money is not a simple power function of the amount of money, nor is the monthly net income the only determinant of emotional responses to imagined gains and losses of money.

The euro, replacing the national currencies in 12 European Union countries, has received considerable attention in the years leading up to the transition. Discussion in the economic literature has focused primarily on the macroeconomic consequences at the European level, such as the effect on inflation rates, economic growth, and levels of employment. Considerably less attention was paid to the anticipated social, cultural, and personal consequences of the single currency. The regular Eurobarometer studies conducted by the European Union have included questions about attitudes toward the euro. However, these opinion surveys do not provide sufficient information about people's underlying hopes, fears, expectations, and values.

Pepermans, Burgoyne, and Müller-Peters (1998) report on a large-scale project conducted in 1997. In all countries of the European Union, data were collected on involvement and knowledge concerning the euro: satisfaction and values, national identity, national pride and European identity, control and expectations, fairness and equity. Pepermans and Verleye (1998) clustered all countries on dimensions such as national economic pride and satisfaction, self-confidence, open-mindedness, and progressive non-nationalistic attitudes. The majority of socio-psychological variables measured in that project had a significant impact on attitudes to the euro. Particular importance attached to knowledge and involvement, life satisfaction and values, national identity and pride, economic expectations, and fairness considerations. Van Everdingen and van Raaij (1998) found macro- and microeconomic expectations affecting attitudes toward the euro. Müller-Peters (1998) concentrates on national identity and its impact on attitudes to the euro. National identity was seen either as a dimension of pure categorization, resulting in patriotism, or as a dimension of discrimination, resulting in nationalism. This distinction of European and national patriotism, on the one hand, and the nationalistic stance, on the other, had particular explanatory force. The former fostered a positive attitude toward the euro, while the latter had a negative impact. Similar results were found by Kokkinaki (1998) and Luna-Arocas, Guzmán, Quintanilla, and Farhangmehr (2001). In the UK, which was not introducing the euro at that time, Routh and Burgoyne (1998) found two kinds of attachment to national identity,

cultural and instrumental attachment, each having both direct and indirect influence upon anti-euro sentiment. It was found that only cultural attachment had a direct, amplifying effect upon anti-euro sentiment. Meier and Kirchler (1998) studied the emotional and cognitive roots of attitudes toward the euro. While people opposing the new currency were found to argue mainly on an emotional basis against the euro, people supporting the replacement of national currencies argued in terms of economic, political, and private advantages. Indifferent or neutral attitudes were mainly held by people who claimed not to be properly informed about the procedure of replacement of national currencies and the consequences of the euro.

TAXES AND TAX EVASION

Tax behaviour as a direct link between individual and state has received special attention in economic psychology. In general, taxation is rejected by the population at large, although the state is expected to make public goods available. The reasons for the rejection of taxes include the view that there is little transparency on spending and that politicians introduce taxes for their own ends. Indeed, Ashworth and Heyndels (2000) report that politicians care about electoral, ideological, and self-esteem motives when defining the level of tax burden in their jurisdiction.

The complexity of tax legislation and lack of transparency in the use of funds is decisive for the rejection of taxation. Complexity has been cited as the most serious problem currently faced by taxpayers (Oveson, 2000), and people with less knowledge about tax law perceive the tax system less fair than knowledgeable people (Eriksen & Fallan, 1996). One significant cause of complexity is the desire on the part of policymakers to determine more accurately and equitably taxpayers' relative abilities to pay. Complexity may also result from attempts to prevent abuse and exploitation of the law, and it could be argued that tax complexity and equity should be positively related. Conversely, even complexity intended to determine taxpayers' abilities to pay more accurately and allocate the tax burden will impose additional compliance costs and administrative costs on taxpayers. In some cases these additional costs, the distribution of these costs, and the resulting inefficiencies may actually increase the welfare cost and inequity of the system. Complexity may also result in taxpayer frustration and increased perceptions of inequity independent of any net effect it may have on actual after-tax income distributions (Cuccia & Carnes, 2001). Carnes and Cuccia (1996) report that the negative relation between complexity and equity ratings of specific tax items weakened as the perceived justification for the complexity increased. Most research investigating the relation between tax equity perceptions and compliance is based, either explicitly or implicitly, on equity theory (Adams, 1965). Equity theory posits that people normatively

expect a comparable rate of inputs and outcomes across all parties to an exchange (e.g., exchange equity between the taxpayer and the government and equity across taxpayers), and will be motivated to alter the distribution if a comparable rate is not perceived to exist. Carnes and Cuccia (1996) found that providing explicit justification mitigates the deleterious effect of tax complexity on tax equity judgements.

People try to avoid taxes mainly because they are guided by self-interest. At the core of most economic analyses of tax evasion is the assumption that people avoid tax payments when it is worthwhile to do so. If the perceived benefits of evasion outweigh the perceived costs, then, if it is possible, individuals will evade taxes. Economic standard theory thus suggests that policy variables such as penalty rate, detection probability, and tax rate are important variables. Psychological approaches stress the importance of values, attitudes, norms and morals, and fiscal consciousness. While economic theory suggests that people are outcome-maximizing or optimizing, psychology emphasizes the process that is involved rather than just outcomes (Cullis & Lewis, 1997). Would people be predominantly outcome-oriented, then virtually everybody should evade taxes given the actual punishment rates and the probability of detection (Smith & Kinsey, 1987).

Hessing and Elffers (1985) and Weigel, Hessing, and Elffers (1987) treat tax evasion as defective behaviour within a social dilemma. In social dilemmas, people are faced with a conflict between the pursuit of their own individual outcome and the pursuit of collective outcome. Non-compliance implies individual gain at some cost to others, while compliance may imply gain to others at some cost to oneself. Thus, the tax system presents people with a choice between co-operative behaviour and defective behaviour. This model has been explored in a number of studies and so far has stood up reasonably well. Elffers, Weigel, and Hessing (1987) found, for example, that measures of personal constraint such as fear of punishment, social controls, relevant tax attitudes, etc. did correlate with self-reported evasion but not with officially documented evasion. Conversely, there was a correlation of personal instigation measures such as dissatisfaction with the tax authorities, alienation, competitiveness, etc. with officially documented evasion but not self-reported evasion. Webley, Cole, and Eidjar (2001) tested the model of taxpaying behaviour, asking non-evaders, people who agreed that they might evade tax but did not report ever having done so, and self-reported evaders, about exchange relationship with the government, and other variables. The most important predictor of self-reported tax evasion was perceived opportunity to evade. The next most important was the perceived prevalence of evasion among friends and colleagues. Other determinants of tax compliance were attitudes to tax authorities, egoism, the perceived exchange relationship with government, attitudes to tax evasion, penalty if caught, and horizontal equity.

Empirical results on self-reported tax morality and tax evasion also indi-

cate that tax compliance is influenced by gender. Spicer and Hero (1985) point out that men are less compliant than women, whereas the findings of other researchers suggest the opposite to be true (Friedland, Maital, & Rutenberg, 1978). There is also some evidence indicating that tax morality is likewise affected by attitudes toward the tax system, perceived justice of the tax system, and knowledge of the legal principles underlying tax law (Groenland & van Veldhoven, 1983; Kirchler, 1997; Vogel, 1974; Webley, Robben, Elffers, & Hessing, 1991).

Elffers and Hessing (1997), with reference to prospect theory (Kahneman & Tversky, 1979), advance the proposition that deliberate overwithholding of income taxes will further the tendency to comply. Moreover, it is demonstrated that offering the taxpayer a choice between full itemized deduction or a considerable, overall standard deduction will enhance compliance as well as considerably reduce the efforts needed by the tax authorities to prevent income tax evasion. Schmidt (2001) provides empirical support for such a proposition. Taxpayers in a balance-due prepayment position were more likely to agree with aggressive advice than taxpayers in a refund position. Deliberate overwithholding of income taxes enhances tax compliance (see also Schepanski & Shearer, 1995).

Kirchler and Maciejovsky (2001) also applied prospect theory to explain tax behaviour. It was hypothesized that tax morality is dependent on gain and loss situations and on the reference point used. Kahneman and Tversky's (1979) approach implicitly suggests that tax-related decisions are based on the expected asset position, whereas Schepanski and Shearer (1995) hold that the current asset position best describes the reference point. Kirchler and Maciejovsky (2001) assumed that individual habits affect which reference point is used (i.e., whether a person uses expected asset position or current asset position as a reference point in making tax-related decisions). Self-employed people, who have the option of choosing the cash receipts and disbursements method, were assumed to employ the current asset position in tax-reporting decisions. Therefore, it was predicted that unexpected payments should lead to low tax morality, whereas unexpected refunds should lead to high tax morality. Conversely, business entrepreneurs, who are obliged to use the more restrictive accrual method, think long term and strategically. Thus, the reference point they should employ in making tax-related decisions is their expected asset position. It was predicted and confirmed that expected payments lead to low tax morality, whereas expected refunds lead to high tax morality for this group of respondents.

With regard to tax compliance, tax practitioners play an important role, since most taxpayers rely on their advice. Tan (1999) reports that they assist the government to enforce tax law when it is unambiguous, but assist taxpayers to exploit tax law when it is ambiguous. Tax practitioners, however, assert that it is the taxpayers who insist on aggressive tax reporting. Tan (1999) found that taxpayers, mainly small business owners, agree with the

advice, conservative or aggressive, given by their practitioner. It appears that the practitioners' advice is generally accepted as correct by their clients who are unfamiliar with tax law. Therefore, the literature suggesting taxpayers to be the instigators of aggressive reporting is not strongly supported. Rather, the majority appear to be cautious taxpayers, primarily interested in filing a correct tax return and avoiding serious tax penalties. In addition, the absence of a significant effect of probability of audit and severity of penalties on taxpayer's decisions indicates that tax decisions are not always based on the economic approach of 'utility maximization'. With regard to agreement with aggressive advice, Schmidt (2001) found that agreement increases if given by certified public accountants than by non-certified public accountants.

GOVERNMENT AND POLICY

'Economics is the study of how people and society end up choosing, with or without the use of money, to employ scarce productive resources ...' (Samuelson, 1976, p. 3). A central concern of economics is how scarce goods are allocated by the interaction of supply and demand in a market system. Kemp (1996, 1998a, 1998b) and Kemp and Bolle (1999) studied preferences for distributing goods by the market or government. If a sudden scarcity of a product develops, for unforeseen reasons and through no fault of the supplier or potential customers, should the product be distributed via the market or a system of regulation? Kemp (1996) presented respondents with scenarios where the shortage was brought about accidentally and only half as much of the commodity as was needed or desired was available. The shortages were of French champagne, heating fuel, sports fields, or a drug needed for treating a possibly fatal disease. The market system was not always regarded as the best way to distribute scarce goods. People's preferences for distribution by market or regulation were substantially affected by the nature of the scarcity. In particular, these preferences were strongly influenced by whether or not people's health was at stake, by the number of people affected by the scarcity, by the expected duration of the scarcity, by whether someone can profit substantially from the scarcity, and by whether the supplier or producer is a monopoly.

One relevant question relates to the availability of public goods, their value, and their use for selfish purposes. Yaniv (1997) investigated welfare fraud and welfare stigma and found that stigma constitutes a stronger deterrent to participation than the expected punishment for dishonest claiming. This result is in line with sociologists' and psychologists' contention that the threat of informal sanctions could have larger effects than legal sanctions. Many public goods have no explicit market price. When the value of public goods as well as environmental goods is estimated, frequently a hypothetical or contingent market is created and willingness to pay for them is assessed.

Contingent valuation requires that individuals are asked their willingness to pay or willingness to accept compensation for goods. This method is widely used but not without shortcomings (e.g., Chilton and Hutchinson, 2000; Knetsch, 1994; Morrison, 2000; Posavac, 1998; Ryan & San Miguel, 2000; Svedsäter, 2000).

Governments can introduce changes in tax and interest rates, and by such interventions control economic behaviour to a certain extent. Usually, reactions to such interventions are slow. East and Hogg (2000) put forward the proposition that the government should use advertising to enhance consumers' responsiveness to the marketplace. They argue that if consumers are prompted to be more alert to price and quality differences in the products and services on offer and if they are encouraged to express their complaints to suppliers and to search for alternative products, then competition in the industry will increase. Such increase in competition will ultimately increase the rate of economic growth and lead to positive outcomes, such as lower prices and improved quality.

Finally, the question arises as to how far the state, the market and the economy in general contribute to the satisfaction of needs and the improvement of life satisfaction. Economic growth is, after all, the motive force driving optimal use of resources for the satisfaction of needs and the consequent increase in satisfaction. However, Easterlin (2001) draws a picture in which the delusion of economic growth leads us to a treadmill in which all our efforts bring us no further than our starting point.

ACKNOWLEDGEMENT

The authors are grateful to Gerlinde Fellner and Boris Maciejovsky who critically read a former draft of the review and made valuable suggestions for improvement.

REFERENCES

Abouchedid, K., & Nasser, R. (2001). Poverty attitudes and their determinants in Lebanon's plural society. *Journal of Economic Psychology*, **22**, 271–282.

Adams, J. (1965). Inequity in social exchange. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology* (pp. 267–299). New York: Academic Press.

Albert, M., Güth, W., Kirchler, E., & Maciejovsky, B. (2002). Holistic experimentation versus decomposition: An ultimatum experiment. *Journal of Economic Behavior and Organization*, **48**, 445–453.

Anand, P. (2001). Procedural fairness in economic and social choice: Evidence from a survey of voters. *Journal of Economic Psychology*, **22**, 247–270.

Anderhub, V., Güth, W., Müller, W., & Strobel, M. (2000). An experimental analysis of intertemporal allocation behavior. *Experimental Economics*, **3**, 137–152.

Anderson, B. F., & Settle, J. W. (1996). The influence of portfolio characteristics and investment period on investment choice. *Journal of Economic Psychology*, **17**, 343–358.

Anderson, M. A., & Goldsmith, A. H. (1997). Mr. Keynes' theory of investment: Do forward looking expectations and weight really matter? *Journal of Economic Psychology*, **18**, 547–573.

Antonides, G. (1991). *Psychology in Economics and Business*. Dordrecht: Kluwer.

Antonides, G., Farago, K., Ranyard, R., & Tyszka, T. (1994). Perceptions of economic activities: A cross-country comparison. Paper given at the *IAREP/SABE Conference, Rotterdam, The Netherlands, 10–13 July*.

Aquino, C., Berti, A. E., & Consolati, G. (1996). La ristrutturazione della conoscenza in un sottodominio economico: La banca. *Giornale Italiano di Psicologia*, **23**, 63–84 [in Italian].

Ashworth, J., & Heyndels, B. (2000). A schema-theoretic approach to politicians' definitions of tax issues. *Journal of Economic Psychology*, **21**, 21–41.

Augier, M. (2001). Sublime Simon: The consistent vision of economic psychology's Nobel laureate. *Journal of Economic Psychology*, **22**, 307–334.

Aukutsionek, S. P., & Belianin, A. V. (2001). Quality of forecasts and business performance: A survey study of Russian managers. *Journal of Economic Psychology*, **22**, 661–692.

Bailey, J. R. (1997). Need for cognition and response mode in the active construction of an information domain. *Journal of Economic Psychology*, **18**, 69–85.

Berti, A. E., & Bombi, A. S. (1981). *Il Mondo Economico nel Bambino*. Florence: La Nuova Italia [in Italian].

Bethwaite, J., & Tompkinson, P. (1996). The ultimatum game and non-selfish utility functions. *Journal of Economic Psychology*, **17**, 259–271.

Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, **19**, 377–395.

Bonn, M., Earle, D., Lea, S., & Webley, P. (1999). South African children's views of wealth, poverty, inequality and unemployment. *Journal of Economic Psychology*, **20**, 593–612.

Boone, C., de Brabander, B., & van Witteloostuijn, A. (1999). The impact of personality on behavior in five Prisoner's Dilemma games. *Journal of Economic Psychology*, **20**, 343–377.

Bosman, C. M., Pfann, G. A., Biddle, J. E., & Hamermesh, D. S. (1997). *Beauty Success and Businesses' Beauty Capital* (Working Paper Series No. 6083). Cambridge: National Bureau of Economic Research.

Brandouy, O. (2001). Laboratory incentive structure and control-test design in an experimental asset market. *Journal of Economic Psychology*, **22**, 1–26.

Brandstätter, E., & Brandstätter, H. (1996). What's money worth? Determinants of the subjective value of money. *Journal of Economic Psychology*, **17**, 443–464.

Brandstätter, H. (1997). Becoming an entrepreneur—a question of personality structure? *Journal of Economic Psychology*, **18**(2–3), 157–177.

Brandstätter, H., & Güth, W. (2000). A psychological approach to individual differences in intertemporal consumption patterns. *Journal of Economic Psychology*, **21**, 465–479.

Burlando, R. M. (2000). Values, ethics and economics. *World Futures*, **56**, 241–261.

Butler, A., & Highhouse, S. (2000). Deciding to sell: The effect of prior inaction and offer source. *Journal of Economic Psychology*, **21**, 223–232.

Cairns, J., & van der Pol, M. (2000). Valuing future private and social benefits: The discounted utility model versus hyperbolic discounting models. *Journal of Economic Psychology*, **21**, 191–205.

Callahan, C., & Elliott, C. S. (1996). Listening: A narrative approach to everyday understandings and behavior. *Journal of Economic Psychology*, **17**, 79–114.

Carnes, G., & Cuccia, A. (1996). An analysis of the effect of tax complexity and its perceived justification on equity judgments. *The Journal of the American Taxation Association*, **18**, 40–56.

Champlin, F. C., & Kopelman, R. E. (1991). Hinrichs revisited: Evaluations of income increases. *The Journal of Psychology*, **125**, 359–373.

Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology*, **22**, 605–616.

Chilton, S. M., & Hutchinson, W. G. (1999). Do focus groups contribute anything to the contingent valuation process? *Journal of Economic Psychology*, **20**, 465–483.

Chilton, S. M., & Hutchinson, W. G. (2000). A note on the warm glow of giving and scope sensitivity in contingent valuation studies. *Journal of Economic Psychology*, **21**, 343–349.

Christopher, A. N., & Schlenker, B. R. (2000). The impact of perceived material wealth and perceiver personality on first impressions. *Journal of Economic Psychology*, **21**, 1–19.

Church, B. K., & Zhang, P. (1999). Bargaining behavior and payoff uncertainty: Experimental evidence. *Journal of Economic Psychology*, **20**, 407–429.

Clark, J. M. (1918). Economics and modern psychology. *The Journal of Political Economy*, **26**, 1–30.

Connor, R. A. (1996). More than risk reduction: The investment appeal of insurance. *Journal of Economic Psychology*, **17**, 39–54.

Cram, F., & Ng, S. H. (1994). Children's understanding of public ownership. *European Journal of Social Psychology*, **24**, 469–480.

Cuccia, A. D., & Carnes, G. A. (2001). A closer look at the relation between tax complexity and tax equity perceptions. *Journal of Economic Psychology*, **22**, 113–140.

Cullis, J. G., & Lewis, A. (1997). Why people pay taxes: From a conventional economic model to a model of social convention. *Journal of Economic Psychology*, **18**, 305–321.

Darity, W., & Goldsmith, A. H. (1993). Unemployment, social psychology, and unemployment hysteresis. *Journal of Post Keynesian Economics*, **16**, 55–72.

Davidsson, P., & Wiklund, J. (1997). Values, beliefs, and regional variations in new firm formation rates. *Journal of Economic Psychology*, **18**, 179–199.

Davis, D. D., & Holt, C. A. (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.

De Ruyter, K., & Wetzels, M. (2000). With a little help from my fans—Extending models of pro-social behaviour to explain supporters' intentions to buy soccer club shares. *Journal of Economic Psychology*, **21**, 387–409.

Dittmar, H., & Drury, J. (2000). Self image—is it in the bag? A qualitative comparison between 'ordinary' and 'excessive' consumers. *Journal of Economic Psychology*, **21**, 109–142.

Donkers, B., & van Soest, A. (1999). Subjective measures of household preferences and financial decisions. *Journal of Economic Psychology*, **20**, 613–642.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, **22**, 91–101.

Earl, P. E. (ed.) (2001). *The Legacy of Herbert Simon in Economic Analysis*. Cheltenham, UK: Edward Elgar.

Earl, P. E., & Kemp, S. (eds) (1999). *The Elgar Companion to Consumer Research and Economic Psychology*. Cheltenham, UK: Edward Elgar.

East, R., & Hogg, A. (2000). Advertising for economic change. *Journal of Economic Psychology*, **21**, 577–590.

Easterlin, R. A. (ed.) (2001). *Happiness in Economics*. Cheltenham, UK: Edward Elgar.

Elffers, H., & Hessing, D. J. (1997). Influencing the prospects of tax evasion. *Journal of Economic Psychology*, **18**, 289–304.

Elffers, H. Weigel, R. H., & Hessing, D. J. (1987). The consequences of different strategies of measuring tax evasion behavior. *Journal of Economic Psychology*, **8**, 311–337.

Elmslie, B., & Sedo, S. (1996). Discrimination, social psychology, and hysteresis in labor markets. *Journal of Economic Psychology*, **17**, 465–478.

El-Sehity, T., Haumer, H., Helmenstein, C., Kirchler, E., & Maciejovsky, B. (Forthcoming). Hindsight bias and individual risk attitude within the context of experimental asset markets.

Eriksen, K., & Fallan, L. (1996). Tax knowledge and attitudes towards taxation: A report on a quasi-experiment. *Journal of Economic Psychology*, **17**, 387–402.

Etzioni, A. (1988). *The Moral Dimension. Toward a New Economics*. New York: Free Press.

Falk, A., Gächter, S., & Kovács, J. (1999). Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology*, **20**, 251–284.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, **49**, 283–306.

Fehr, E., & Gächter, S. (1998). Reciprocity and economics: The economic implications of Homo reciprocans. *European Economic Review*, **42**, 845–859.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, **65**, 833–860.

Fehr, E., Kirchler, E., Weichbold, A., & Gächter, S. (1998). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics*, **16**, 324–351.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, **114**, 817–868.

Ferrari, L., & Romano, D. (1999). *Mente e Denaro. Introduzione alla Psicologia Economica* [Mind and Money. Introduction to Economic Psychology]. Milan: R. Cortina [in Italian].

Fershtman, C. (1996). On the value of incumbency: Managerial reference points and loss aversion. *Journal of Economic Psychology*, **17**, 245–258.

Frank, H., & Korunka, C. (1996). Zum Informations- und Entscheidungsverhalten von Unternehmensgründern. Der Zusammenhang von 'Handlungskontrolle' und Gründungserfolg. *Zeitschrift für Betriebswirtschaft*, **66**, 947–963 [in German].

Frank, R. H. (1985). *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.

Frey, B. S. (1990). Entscheidungsanomalien: Die Sicht der Ökonomie. *Psychologische Rundschau*, **41**, 67–83 [in German].

Friedland, N., Maital, S., & Rutenberg, A. (1978). A simulation study of income tax evasion. *Journal of Public Economics*, **10**, 107–116.

Frijters, P. (2000). Do individuals try to maximize general satisfaction? *Journal of Economic Psychology*, **21**, 281–304.

Furnham, A. (1997). The relationship between work and economic values. *Journal of Economic Psychology*, **18**, 1–14.
Furnham, A. (1999). The saving and spending habits of young people. *Journal of Economic Psychology*, **20**, 677–697.
Furnham, A. (2001). Parental attitudes to pocket money/allowances for children. *Journal of Economic Psychology*, **22**, 397–422.
Furnham, A., & Lewis, A. (1986). *Economic Mind. The Social Psychology of Economic Behaviour*. Brighton, UK: Wheatsheaf Books.
Gerke, W., Arneth, S., & Syha, C. (2000). The impact of the order book privilege on traders' behavior and the market process: An experimental study. *Journal of Economic Psychology*, **21**, 167–189.
Gigerenzer, G., Todd, P. M., & The ABC Research Group (1999). *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.
Gneezy, U., Güth, W., & Verboven, F. (2000). Presents or investments? An experimental analysis. *Journal of Economic Psychology*, **21**, 481–493.
Goldsmith, A. H., Veum, J. R., & Darity, W. Jr (1996). The impact of labor force history on self-esteem and its component parts, anxiety, alienation and depression. *Journal of Economic Psychology*, **17**, 183–220.
Goldsmith, A. H., Veum, J. R., & Darity, W. Jr (2000). Working hard for the money? Efficiency wages and worker effort. *Journal of Economic Psychology*, **21**, 351–385.
Gravelle, H., & Rees, R. (1981). *Microeconomics*. London: Longman.
Groenland, E. A. G., Kuylen, A. A. A., & Bloem, J. G. (1996). The dynamic process tracing approach: Towards the development of a new methodology for the analysis of complex consumer behaviours. *Journal of Economic Psychology*, **17**, 809–826.
Groenland, E. A. G., & van Veldhoven, G. M. (1983). Tax evasion behavior: A psychological framework. *Journal of Economic Psychology,* **3**, 129–144.
Groot, W., & van den Brink, H. M. (1999). The price of stress. *Journal of Economic Psychology*, **20**, 83–103.
Gunnarsson, J., & Wahlund, R. (1995). *Hushallens Finansiella Strategier*. Stockholm: The Economic Research Institute at the Stockholm School of Economics [in Swedish].
Gunnarsson, J., & Wahlund, R. (1997). Household financial strategies in Sweden: An exploratory study. *Journal of Economic Psychology*, **18**, 201–233.
Güth, W. (2000). Boundedly rational decision emergence—a general perspective and some selective illustrations. *Journal of Economic Psychology*, **21**, 433–458.
Güth, W., Krahnen, J. P., & Rieck, C. (1997). Financial markets with asymmetric information: A pilot study focusing on insider advantages. *Journal of Economic Psychology*, **18**, 235–257.
Güth, W., Ockenfels, P., & Wendel, M. (1997). Cooperation based on trust. An experimental investigation. *Journal of Economic Psychology*, **18**, 15–43.
Güth, W., Schmittberger, R., & Schwarzer, B. (1982). An experimental study of ultimatum bargaining. *Journal of Economic Behavior and Organization*, **3**, 367–388.
Hayhoe, C. R., Leach, L., & Turner, P. R. (1999). Discriminating the number of credit cards held by college students using credit and money attitudes. *Journal of Economic Psychology*, **20**, 643–656.
Hessing, D. J., & Elffers, H. (1985). Economic man or social man? A social orientation model for individual behaviour in social dilemmas. In H. Brandstätter, & E. Kirchler (eds), *Economic Psychology* (pp. 195–203). Linz, Austria: Trauner.
Hey, J. D. (1991). *Experiments in Economics*. Oxford: Basil Blackwell.

Hinrichs, J. R. (1969). Correlates of employee evaluations of pay increases. *Journal of Applied Psychology*, **53**, 481–489.

Hoorens, V., Remmers, N., & van de Riet, K. (1999). Time is an amazingly variable amount of money: Endowment and ownership effects in the subjective value of working time. *Journal of Economic Psychology*, **20**, 383–405.

Huck, S. (1999). Responder behavior in ultimatum offer games with incomplete information. *Journal of Economic Psychology*, **20**, 183–206.

Huck, S., & Oechssler, J. (2000). Informational cascades in the laboratory: Do they occur for the right reasons? *Journal of Economic Psychology*, **21**, 661–671.

Huck, S., & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment. *Journal of Economic Psychology*, **20**, 699–715.

Hunt, M. (1996). The individual, society, or both? A comparison of Black, Latino and White beliefs about the causes of poverty. *Social Forces*, **75**, 293–322.

Jain, A. K., & Joy, A. (1997). Money matters: An exploratory study of the socio-cultural context of consumption, saving, and investment patterns. *Journal of Economic Psychology*, **18**, 649–675.

Janssen, M. A., & Jager, W. (2001). Fashions, habits and changing preferences: Simulation of psychological factors affecting market dynamics. *Journal of Economic Psychology*, **22**, 745–772.

Jungermann, H., Pfister, H.-R., & Fischer, K. (1998). *Die Psychologie der Entscheidung*. Heidelberg, Germany: Spektrum [in German].

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, **76**, 728–741.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.

Kasser, T., & Grow Kasser, V. (2001). The dreams of people high and low in materialism. *Journal of Economic Psychology*, **22**, 693–719.

Katona, G. (1951). *Psychological Analysis of Economic Behavior*. New York: Elsevier.

Katona, G. (1975) *Psychological Economics*. New York: Elsevier.

Kemp, S. (1996). Preferences for distributing goods in times of shortage. *Journal of Economic Psychology*, **17**, 615–628.

Kemp, S. (1998a). Rating the values of government and market supplied goods. *Journal of Economic Psychology*, **19**, 447–461.

Kemp, S. (1998b). Perceiving luxury and necessity. *Journal of Economic Psychology*, **19**, 591–606.

Kemp, S., & Bolle, F. (1999). Preferences in distributing scarce goods. *Journal of Economic Psychology*, **20**, 105–120.

Kemp, S., & Willetts, K. (1996). Remembering the price of wool. *Journal of Economic Psychology*, **17**, 115–126.

Khalil, E. L. (1996). Respect, admiration, aggrandizement: Adam Smith as economic psychologist. *Journal of Economic Psychology*, **17**, 555–578.

Kirchler, E. (1997). Balance between giving and receiving: Tax morality and satisfaction with fiscal policy as they relate to the perceived just distribution of public resources. *Reitaku International Journal of Economic Studies*, **5**, 59–70.

Kirchler, E. (1998). Differential representations of taxes: Analysis of free associations and judgments of five employment groups. *Journal of Socio Economics*, **27**, 117-131.

Kirchler, E. (1999). *Wirtschaftspsychologie*. Göttingen, Germany: Hogrefe [in German].

Kirchler, E., Fehr, E., & Evans, R. (1996). Social exchange in the labor market: Reciprocity and trust versus egoistic money maximization. *Journal of Economic Psychology*, **17**, 313–342.
Kirchler, E., & Maciejovsky, B. (2001). Tax compliance within the context of gain and loss situations, expected and current asset position, and profession. *Journal of Economic Psychology*, **22**, 173–194.
Kirchler, E., & Maciejovsky, B. (2002). Simultaneous over- and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty*, **25**, 65–85.
Kirchler, E., Rodler, C., Hölzl, E., & Meier, K. (2001). *Conflict and Decision Making in Close Relationships*. Hove, UK: Psychology Press.
Knetsch, J. L. (1994). Environmental valuation: Some problems of wrong questions and misleading answers. *Environmental Values*, **3**, 351–368.
Kokkinaki, F. (1998). Attitudes towards European Monetary Union in Greece: Antecedents, strength and consequences. *Journal of Economic Psychology*, **19**, 775–796.
Korunka, C., Frank, H., & Becker, P. (1993). Persönlichkeitseigenschaften von Unternehmensgründern. Erfolgsfaktor oder vernachlässigbare Restgröße in der Theorie der Unternehmensgründung. *Internationales Gewerbearchiv Zeitschrift für Klein- und Mittelunternehmen*, **41**, 169–188 [in German].
Krahnen, J. P., Rieck, C., & Theissen, E. (1997). Inferring risk attitudes from certainty equivalents: Some lessons from an experimental study. *Journal of Economic Psychology*, **18**, 469–486.
Kristensen, H., & Gärling, T. (1997). Determinants of buyers' aspiration and reservation price. *Journal of Economic Psychology*, **18**, 487–503.
Kühberger, A. (1995). The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes*, **62**, 230–240.
Kyle, D. M., & Mahler, H. I. M. (1996). The effect of hair color and cosmetic use on perceptions of a female's ability. *Psychology of Women Quarterly*, **20**, 447–455.
Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, **62**, 443–477.
Lea, S. E. G., Tarpy, R. M., & Webley, P. (1987). *The Individual in the Economy: A Survey of Economic Psychology*. Cambridge: Cambridge University Press.
Li, S. (1996). An additional violation of transitivity and independence between alternatives. *Journal of Economic Psychology*, **17**, 645–650.
Li, S. (1998). Can the conditions governing the framing effect be determined? *Journal of Economic Psychology*, **19**, 133–153.
Lim, V. K. G., & Teo, T. S. H. (1997). Sex, money and financial hardship: An empirical study of attitudes towards money among undergraduates in Singapore. *Journal of Economic Psychology*, **18**, 369–386.
Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, **107**, 573–597.
Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, **92**, 805–824.
Luna-Arocas, R., Guzmán, G., Quintanilla, I., & Farhangmehr, M. (2001). The euro and European identity: The Spanish and Portuguese case. *Journal of Economic Psychology*, **22**, 441–460.
Lundberg, C. G. (2000). Made sense and remembered sense: Sensemaking through abduction. *Journal of Economic Psychology*, **21**, 691–709.
Lunt, P. (1996). Discourses of savings. *Journal of Economic Psychology*, **17**, 677–690.
Mackenzie, C. (1997). Where are the motives? A problem with evidence in the work of Richard Thaler. *Journal of Economic Psychology*, **18**, 123–135.

Mador, G., Sonsino, D., & Benzion, U. (2000). On complexity and lotteries' evaluation—three experimental observations. *Journal of Economic Psychology*, **21**, 625–637.

Meier, K., & Kirchler, E. (1998). Social representations of the euro in Austria. *Journal of Economic Psychology*, **19**, 755–774.

Meier, K., Kirchler, E., & Hubert, A.-C. (1999). Savings and investment decisions within private households: Spouses' dominance in decisions on various forms of investment. *Journal of Economic Psychology*, **20**, 499–519.

Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (1998). Judgment and decision making. *Annual Review of Psychology*, **49**, 447–477.

Mitchell, W. C. (1914). Human behavior and economics: A survey of recent literature. *The Quarterly Journal of Economics*, **29**, 1–47.

Mitropoulos, A. (2001). Learning under minimal information: An experiment on mutual fate control. *Journal of Economic Psychology*, **22**, 523–558.

Morrison, G. C. (2000). WTP and WTA in repeated trial experiments: Learning or leading? *Journal of Economic Psychology*, **21**, 57–72.

Müller, W. (2001). Strategies, heuristics, and the relevance of risk-aversion in a dynamic decision problem. *Journal of Economic Psychology*, **22**, 493–522.

Müller-Peters, A. (1998). The significance of national pride and national identity to the attitude toward the single European currency: A Europe-wide comparison. *Journal of Economic Psychology*, **19**, 701–719.

Münsterberg, H. (1912). *Psychologie und Wirtschaftsleben*. Ein Beitrag zur angewandten Experimentalpsychologie. Leipzig, Germany: Barth [in German].

Murnighan, J. K., & Saxon, M. S. (1998). Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, **19**, 415–445.

Ouwersloot, H., Nijkamp, P., & Rietveld, P. (1998). Errors in probability updating behaviour: Measurement and impact analysis. *Journal of Economic Psychology*, **19**, 535–563.

Overton, A. A., & MacFadyen, A. J. (1998). Time discounting and the estimation of loan duration. *Journal of Economic Psychology*, **19**, 607–618.

Oveson, W. V. (2000). *National Taxpayer Advocate's Annual Report to Congress*. Washington, DC: Office of Taxpayer Advocate.

Pålsson, A.-M. (1996). Does the degree of relative risk aversion vary with household characteristics? *Journal of Economic Psychology*, **17**, 771–788.

Pepermans, R., Burgoyne, C. B., & Müller-Peters, A. (1998). Editorial. *Journal of Economic Psychology*, **19**, 657–661.

Pepermans, R., & Verleye, G. (1998). A unified Europe? How euro attitudes relate to psychological differences between countries. *Journal of Economic Psychology*, **19**, 681–699.

Pepermans, R., Verleye, G., & Van Cappellen, S. (1996). 'Wallbanking', innovativeness and computer attitudes: 25–40-year-old ATM-users on the spot. *Journal of Economic Psychology*, **17**, 731–748.

Pingle, M. (1997). Submitting to authority: Its effect on decision-making. *Journal of Economic Psychology*, **18**, 45–68.

Plug, E. J. S. (2001). Season of birth, schooling and earnings. *Journal of Economic Psychology*, **22**, 641–660.

Plug, E. J. S., & van Praag, B. M. S. (1998). Similarity in response behavior between household members: An application to income evaluation. *Journal of Economic Psychology*, **19**, 497–513.

Posavac, S. S. (1998). Strategic overbidding in contingent valuation: Stated economic value of public goods varies according to consumers expectations of funding source. *Journal of Economic Psychology*, **19**, 205–214.

Powell, M., & Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, **18**, 605–628.

Rambo, W. W., & Pinto, J. N. (1989). Employees' perception of pay increases. *Journal of Occupational Psychology*, **62**, 135–145.

Rodgers, W. (1999). The influence of conflicting information on novices and loan officers' actions. *Journal of Economic Psychology*, **20**, 123–145.

Routh, D. A., & Burgoyne, C. B. (1998). Being in two minds about a single currency: A UK perspective on the euro. *Journal of Economic Psychology*, **19**, 741–754.

Rumiati, R., & Lotto, L. (1996). Varieties of money. Experts' and non-experts' typicality judgments. *Journal of Economic Psychology*, **17**, 403–414.

Ryan, M., & San Miguel, F. (2000). Testing for consistency in willingness to pay experiments. *Journal of Economic Psychology*, **21**, 305–317.

Sacco, P. L. (1996). Subjective metaphysics and learning from experience: The causal psychology of rational choice. *Journal of Economic Psychology*, **17**, 221–244.

Samuelson, P. A. (1976). *Economics*. New York: McGraw-Hill.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, **22**, 617–640.

Schaubroeck, J. (1996). Pay status hierarchy and organizational attachment. *Journal of Economic Psychology*, **17**, 579–590.

Schepanski, A., & Shearer, T. (1995). A prospect theory account of the income tax withholding phenomenon. *Organizational Behavior and Human Decision Making Processes*, **63**, 174–186.

Schmidt, D. R. (2001). The prospects of taxpayer agreement with aggressive tax advice. *Journal of Economic Psychology*, **22**, 157–172.

Schram, A., & Sonnemans, J. (1996). Why people vote: Experimental evidence. *Journal of Economic Psychology*, **17**, 417–442.

Schram, A., & van Winden, F. A. A. M. (1991). Why people vote: Free riding and the production and consumption of social pressure. *Journal of Economic Psychology*, **12**, 575–620.

Schuldner, K. (2001). Geschichte und Inhalte der Ökonomischen Psychologie. Unpublished thesis, University of Vienna [in German].

Schwer, R. K., & Daneshvary, R. (2000). Keeping up one's appearance: Its importance and the choice of type of hair-grooming establishment. *Journal of Economic Psychology*, **21**, 207–222.

Seligman, D. A., & Schwartz, B. (1997). Domain specificity of fairness judgments in economic transactions. *Journal of Economic Psychology*, **18**, 579–604.

Sen, A. (1993). Markets and freedoms. *Oxford Economic Papers*, **45**, 519–541.

Shavit, T., Sonsino, D., & Benzion, U. (2001). A comparative study of lotteries—evaluation in class and on the Web. *Journal of Economic Psychology*, **22**, 483–492.

Shefrin, H. (ed.) (2001). *Behavioral Finance*. (Vols 1–3). Cheltenham, UK: Edward Elgar.

Simon, H. (1957). *Models of Man*. New York: John Wiley & Sons.

Simon, H. (1982). *Models of Bounded Rationality*. Cambridge, MA: MIT Press.

Singh, L. B., Singh, A. K., & Rani, A. (1996). Level of self-concepts in educated unemployed young men in India: An empirical analysis. *Journal of Economic Psychology*, **17**, 629–644.

Smith, A. (1776). *An Enquiry into the Nature and Causes of the Wealth of Nations*. London: Bell.

Smith, K. W., & Kinsey, K. A. (1987). Understanding tax-paying behavior: A conceptual framework with implications for research. *Law and Society Review*, **21**, 639–663.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, **66**, 274–277.

Snelders, H. M. J., Hussein, G., Lea, S. E. G., & Webley, P. (1992). The polymorphous concept of money. *Journal of Economic Psychology*, **13**, 71–92.

Sonnegård, J. (1996). Determination of first movers in sequential bargaining games: An experimental study. *Journal of Economic Psychology*, **17**, 359–386.

Sonnemans, J. (2000). Decisions and strategies in a sequential search experiment. *Journal of Economic Psychology*, **21**, 91–102.

Spicer, M. W., & Hero, R. E. (1985). Tax evasion and heuristics: A research note. *Journal of Public Economics*, **26**, 263–267.

Stroeker, N. E., & Antonides, G. (1997). The process of reaching an agreement in second-hand markets for consumer durables. *Journal of Economic Psychology*, **18**, 341–367.

Strümpel, B., & Katona, G. (1983). Psychologie gesamtwirtschaftlicher Prozesse. In M. Irle, & W. Bussmann (eds), *Enzyklopädie der Psychologie. Marktpsychologie als Sozialwissenschaft* (pp. 225–281). Göttingen, Germany: Hogrefe [in German].

Suleiman, R. (1996). Expectations and fairness in a modified Ultimatum game. *Journal of Economic Psychology*, **17**, 531–554.

Supphellen, M., & Nelson, M. R. (2001). Developing, exploring, and validating a typology of private philanthropic decision making. *Journal of Economic Psychology*, **22**, 573–604.

Svedsäter, H. (2000). Contingent valuation of global environmental resources: Test of perfect and regular embedding. *Journal of Economic Psychology*, **21**, 605–623.

Swedberg, R. (1991). Major traditions of economics sociology. *Annual Review of Sociology*, **17**, 251–276.

Tan, L. M. (1999). Taxpayers' preference for type of advice from tax practitioner: A preliminary examination. *Journal of Economic Psychology*, **20**, 431–447.

Tang, T. L.-P. (1996). Pay differentials as a function of rater's sex, money ethic, and job incumbent's sex: A test of the Matthew Effect. *Journal of Economic Psychology*, **17**, 127–144.

Tarde, G. (1902). *La Psychologie Économique*. Paris: Alcan [in French].

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economic Letters*, **8**, 201–207.

Thaler, R. (1992). *The Winner's Curse*. Princeton, NJ: Princeton University Press.

Thaler, R. (1999). The end of behavioral finance. *Financial Analysts Journal*, **55**, 12–19.

Theodoulou, S. (1996). Construing economic and political reality. *Journal of Economic Psychology*, **17**, 499–516.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, **211**, 543–458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**, 297–323.

Tykocinski, O. E., Pittman, T. S., & Tuttle, E. E. (1995). Inaction inertia: Foregoing future benefits as a result of an initial failure to act. *Journal of Personality and Social Psychology*, **68**, 793–803.

Tyszka, T. (1994). Cognitive representations of economics. Paper given at the *IAREP/SABE Conference, Rotterdam, The Netherlands, 10-13 July*.

Tyszka, T., & Sokolowska, J. (1992). Perception and judgments of the economic system. *Journal of Economic Psychology*, **13**, 421–448.

Unser, M. (2000). Lower partial moments as measures of perceived risk: An experimental study. *Journal of Economic Psychology*, **21**, 253–280.

Van der Heijden, E. C. M., Nelissen, J. H. M., Potters, J. J. M., & Verbon, H. A. A. (1998). The poverty game and the pension game: The role of reciprocity. *Journal of Economic Psychology*, **19**, 5–41.

Van Dijk E., & van Knippenberg, D. (1996). Buying and selling exchange goods: Loss aversion and the endowment effect. *Journal of Economic Psychology*, **17**, 517–524.

Van Dijk, E., & van Knippenberg, D. (1998). Trading wine: On the endowment effect, loss aversion, and the comparability of consumer goods. *Journal of Economic Psychology*, **19**, 485–495.

Van Everdingen, Y. M., & van Raaij, W. F. (1998). The Dutch people and the euro: A structural equations analysis relating national identity and economic expectations to attitude towards the euro. *Journal of Economic Psychology*, **19**, 721–740.

Van Raaij, W. F. (1998). The life and work of Amos Tversky. *Journal of Economic Psychology*, **19**, 515–527.

Van Raaij, W. F., van Veldhoven, G. M., & Wärneryd, K.-E. (eds) (1988) *Handbook of Economic Psychology*. Dordrecht, The Netherlands: Kluwer.

Van Veldhoven, G. M. (1988). Dynamic aspects of economic behavior: Some determinants. In W. F. van Raaij, G. M. van Veldhoven, & K. E. Wärneryd (eds), *Handbook of Economic Psychology* (pp. 52–73). Dordrecht, The Netherlands: Kluwer.

Veblen, T. (1899). *The Theory of the Leisure Class*. New York: Macmillan.

Viaud, J., & Roland-Lévy, C. (2000). A positional and representational analysis of consumption: Households when facing debt and credit. *Journal of Economic Psychology*, **21**, 411–432.

Vogel, J. (1974). Taxation and public opinion in Sweden: An interpretation of recent survey data. *National Tax Journal*, **27**, 499–513.

von Rosenstiel, L. & Ewald, G. (1979). *Marktpsychologie* (Vols 1 and 2). Stuttgart, Germany: Kohlhammer [in German].

Wahlund, R., & Gunnarsson, J. (1996). Mental discounting and financial strategies. *Journal of Economic Psychology*, **17**, 709–730.

Wakely, T. M. (1997). The output decision of the firm—A behavioural algorithm. *Journal of Economic Psychology*, **18**, 111–121.

Walker, C. M. (1996). Financial management, coping and debt in households under financial strain. *Journal of Economic Psychology*, **17**, 789–808.

Wang, X. T. (1996). Domain-specific rationality in human choices: Violation of utility axioms and social contexts. *Cognition*, **60**, 31–63.

Wärneryd, K.-E. (1988). The psychology of innovative entrepreneurship. In W. F. van Raaij, G. M. van Veldhoven, & K.-E. Wärneryd (eds), *Handbook of Economic Psychology* (pp. 404–447). Dordrecht, The Netherlands: Kluwer.

Wärneryd, K.-E. (1993). Psychology + economics = economic psychology? Unpublished paper, Stockholm School of Economics.

Wärneryd, K.-E. (1996). Risk attitudes and risky behavior. *Journal of Economic Psychology*, **17**, 749–770.

Wärneryd, K.-E. (1999). *The Psychology of Saving: A Study on Economic Psychology*. Cheltenham, UK: Edward Elgar.

Wärneryd, K.-E. (2001). *Stock-Market Psychology. How People Value and Trade Stocks*. Cheltenham, UK: Edward Elgar.

Wärneryd, K.-E., & Westlund, K. (1992). Ethics and economic affairs in the world of finance. *Journal of Economic Psychology*, **14**, 523–539.

Waters, L. E., & Moore, K. A. (2001). Coping with economic deprivation during unemployment. *Journal of Economic Psychology*, **22**, 461–482.

Wesley, P., Burgoyne, C. B., Lea, S. E. G., & Young, B. M. (2001). *The Economic Psychology of Everyday Life*. Hove, UK: Psychology Press.

Webley, P., Cole, M., & Eidjar, O.-P. (2001). The prediction of self-reported and hypothetical tax-evasion: Evidence from England, France and Norway. *Journal of Economic Psychology*, **22**, 141–156.

Webley, P., Lewis, A., & Mackenzie, C. (2001). Commitment among ethical investors: An experimental approach. *Journal of Economic Psychology*, **22**, 27–42.

Webley, P., Robben, H., Elffers, H., & Hessing, D. (1991). *Tax Evasion. An Experimental Approach*. Cambridge: Cambridge University Press.

Weigel, R. H., Hessing, D. J., & Elffers, H. (1987). Tax evasion research: A critical appraisal and a theoretical model. *Journal of Economic Psychology*, **8**, 215–235.

Williamson, M. R., & Wearing, A. J. (1996). Lay people's cognitive models of the economy. *Journal of Economic Psychology*, **17**, 3–38.

Winnett, A., & Lewis, A. (2000). 'You'd have to be green to invest in this': Popular economic models, financial journalism, and ethical investment. *Journal of Economic Psychology*, **21**, 319–339.

Wiswede, G. (2000). *Einführung in die Wirtschaftspsychologie*. Munich, Germany: UTB [in German].

Yaniv, G. (1997). Welfare fraud and welfare stigma. *Journal of Economic Psychology*, **18**, 435–451.

Zeelenberg, M., & Van Dijk, E. (1997). A reverse sunk cost effect in risky decision making: Sometimes we have too much invested to gamble. *Journal of Economic Psychology*, **18**, 677–691.

Zwick, R., Rapoport, A., & Weg, E. (2000). Invariance failure under subgame perfectness in sequential bargaining. *Journal of Economic Psychology*, **21**, 517–544.

Chapter 3

SLEEPINESS IN THE WORKPLACE: CAUSES, CONSEQUENCES, AND COUNTERMEASURES

Autumn D. Krauss, Peter Y. Chen, and Sarah DeArmond
Colorado State University

and Bill Moorcroft
Sleep and Dreams Laboratory, Luther College

Feeling sleepy during the workday as a result of lacking sufficient sleep has become an epidemic problem in the working population according to annual polls conducted by the National Sleep Foundation (NSF) from 1998 to 2002. The most recent poll (NSF, 2002) showed that 39% of respondents reported getting less than seven hours of sleep on weeknights, which is one hour less than recommended by sleep experts. The poll's findings further suggested a negative relationship between sleep hours and daytime sleepiness, with 37% of respondents reporting daytime sleepiness and 6% the use of medications to stay awake.

It is our contention that sleepiness in the workplace has been a neglected occupational health topic in Industrial and Organizational (I/O) Psychology. The potential consequences pertaining to sleepiness in the workplace have profound practical, health, and legal implications for organizations. According to the 2002 poll, described above, over 90% of respondents believed that their work performance and safety were influenced by their sleep debt. In addition, over 60% of respondents believed that, as a result of sleep debt, they had difficulty reading business documents, taking on additional tasks, making thought-out decisions, or recalling things they had just heard. The series of surveys also revealed that the greater the number of hours worked, the less sleep obtained, the more negative emotions (e.g., anger, anxiety) experienced.

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

Arguably, sleepiness on the job may not only affect employees and their organizations, but also innocent bystanders. For example, the very recent tragedy consisting of a towboat causing the collapse of a 500-ft section of a bridge in Oklahoma may be related to sleepiness on the job. The National Transportation Safety Board investigator stated that the captain of the towboat had less than 10 hours of sleep within the 42 hours preceding the accident (CBS, 2002, May 30), although the captain insisted that he did not fall asleep but merely blacked out minutes before the accident (Romano, 2002, May 30).

In a recent litigation, *Glander et al. vs Peeler et al.* in the Superior Court, County of Sacramento, California No. 98AS03253 (cited in Mitler, 2002), the defendant, who was a nurse working 12-hr night shifts, rear-ended a pick-up truck resulting in the deaths of its driver and a passenger. The family of the deceased sued the nurse and the hospital where the defendant worked. They claimed that the hospital was partially responsible for the deaths of their family members, because they required the nurse to stay an additional two to three hours to attend a skills workshop. The jury found that the nurse was 75% at fault and the hospital was 25% at fault and awarded the plaintiffs approximately $1,300,000, which was entirely paid by the hospital.

Considering the above cases, it is imperative for I/O psychologists to explore plausible causes and consequences of sleepiness in the workplace and to develop subsequent countermeasures to mitigate these consequences. To initiate this attempt, we present our chapter, which consists of seven sections pertaining to sleepiness in the workplace. Initially, we offer a basic review regarding sleepiness in general with particular emphasis on daytime sleepiness. This fundamental information will facilitate comprehension of more specific reviews throughout the chapter. Because sleep disorders are a topic beyond the scope of I/O psychology and people suffering from them should consult with medical professionals, research concerning sleep disorders will not be discussed (refer to Moorcroft, 2003 for information about sleep disorders). Following this basic information, we review common objective and subjective measures of the sleepiness state.

In the next four sections, we focus on possible causes as well as consequences of sleepiness in the workplace at both individual and organizational levels. The general foci of sleep research have been that of sleep disorders, general sleep patterns in various developmental stages, the effects of work schedule on sleep (Garbarino et al., 2002), and the relationship between type of occupation and sleep (e.g., physician, Lewis, Blagrove, & Ebden, 2002; professional driver, Horne & Reyner, 1995). Because the focus of this chapter is sleepiness in the workplace, only limited studies were retrieved from the sleep literature. Therefore, we have broadened our review to include empirical findings from the traditional I/O literature, in which the relationships among sleepiness surrogates (e.g., sleep quality, somatic symptoms), causes (e.g., job stressors, job characteristics), and consequences (e.g., anger at

work, motivation) have been investigated. Specifically, we contend that sleep quality, sleep quantity, sleep disturbances, sleep deprivation, fatigue, and somatic symptoms are all related to the state of sleepiness (Lee, Hicks, & Nino-Murcia, 1991). Since these other sleep variables rather than workplace sleepiness are usually the foci of past research, readers should be cognizant of the inevitable inferential leaps required during these sections of the chapter. Finally, we propose individual and organizational countermeasures based on empirical findings from the sleep, job stress, and job design research domains.

DAYTIME SLEEPINESS

Daytime sleepiness has been a growing problem in the fast-paced workplace that exists today. This phenomenon can be understood from two diverse, but not mutually exclusive, perspectives: physiological and subjective. Physiological daytime sleepiness is simply the result of biological reactions to unfulfilled sleep quotas or disruption of the internal biological clock. In contrast to physiological sleepiness, subjective daytime sleepiness is less obvious and more difficult to research in the conventional sleep literature. It can be influenced by physiological sleepiness but can also be affected by various other plausible factors such as the work environment (e.g., light, noise, temperature), job or task characteristics (e.g., stationary posture, vigilance), stressors and strains, motivation, or diet (e.g., consumption of stimulants). Because subjective sleepiness is influenced by these other factors, the self-perceived level of sleepiness may conceal the actual amount of physiological sleepiness. Consequently, it is not unusual for people to underestimate their physiological level of sleepiness and their sleep need.

Physiological daytime sleepiness is dependent upon the quantity, quality, and timing of prior sleep plus the amount of prior wakefulness. Sleep is a function of the brain (Culebras, 2002), and can be viewed as 'a reversible behavioral state of perceptual disengagement from an unresponsiveness to the environment' (Carskadon & Dement, 2000, p. 15). Sleep is controlled by neural centers that are primarily located in the brainstem, diencephalons, and thalamus. As such, sleep can alter the levels of some bodily components (e.g., body temperature, hormones). Because of the physiological nature of sleep, it would be difficult to train employees to need less amounts of sleep, although it is possible to help them learn to sleep better.

The need to sleep is controlled by two interacting neurobiological components—a sleep quota and an internal 24-hr biological clock, which is also known as circadian rhythm. The sleep quota is determined by two opposing factors—the amount of previous sleep and the amount of time awake. The relationship is such that sleep quota increases with wake time and decreases with sleep time. It takes roughly one hour of sleep to adequately compensate

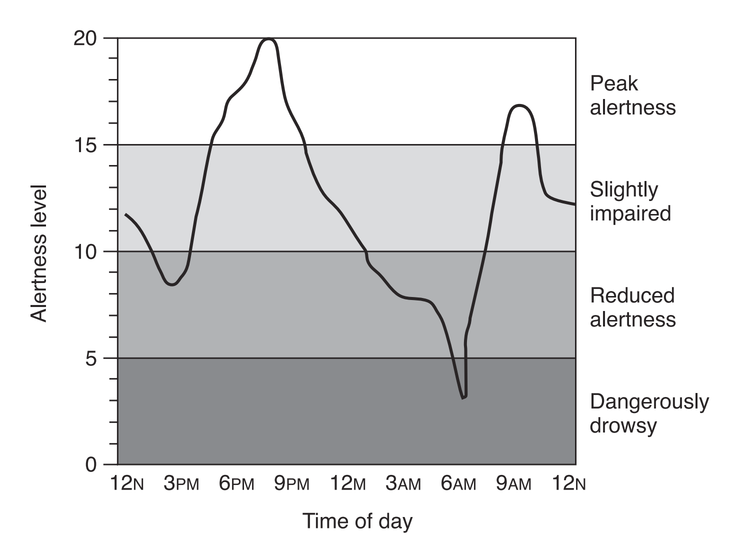


Figure 3.1 Sleep propensity as a function of the 24-hr clock. (*Source:* National Highway Traffic Safety Administration (2002). Sick and tired of waking up sick and tired. Retrieved on May 28, 2002 from http://www.nhtsa.dot.gov/people/injury/drows_driving/wbroch/wbrochure.pdf.)

for two hours of wakefulness in the average person. The internal clock, synchronized with the external rotation of lightness and darkness, is a brain mechanism controlled by the suprachiasmatic nucleus of the hypothalamus (Culebras, 2002). In general, people feel a biological need to sleep at night and, to an important but lesser extent, during the middle of the afternoon, as depicted in Figure 3.1. Generally, people reach peak alertness around 9–10 a.m. and 8–9 p.m., and primary sleepiness occurs between 10 p.m. and 6 a.m. Interestingly, alertness is also noticeably reduced some time between 2 and 4 p.m., which is considered a secondary sleepiness zone. Note that these times are shifted up to an hour earlier or later in some individuals.

Circadian rhythm includes three aspects, phase, rigidity, and vigor (Folkard, Monk, & Lobban, 1979). Phase, often labeled morning–evening orientation, refers to a biological preference for morning or evening activity. Rigidity refers to the stability of the circadian rhythm, while vigor is the amplitude of the circadian rhythm.

Laboratory experiments, in agreement with real-world observations, have shown that the brain functions less efficiently and shifts quickly into sleep if sleep quotas are not met or the internal biological clock is disrupted. With the loss of merely one night of sleep, noticeable reductions in brain activity especially in the prefrontal cortex appeared (Drummond & Brown, 2001). The shift to sleep occurs not only quickly but also sometimes uncontrollably, resulting in sleep that lasts seconds ('microsleeps') or sleep that is sustained

for longer periods of time. Both the inefficiency of brain functioning and uncontrollable sleep could lead to various deleterious consequences.

Note that most employees, with the exception of those who engage in certain types of professions, do not often experience continuous, total sleep deprivation. Instead, it is often the accumulation of insufficient sleep night after night or irregular sleep patterns that may result in sleepiness on the job. Recent research has demonstrated that accumulated sleep debt leads to performance decrements in areas such as attentiveness, psychomotor co-ordination, and information processing (Dinges et al., 1997; Harrison & Horne, 1999; Horne, 1988). Metaphorically, the effects of sleepiness on performance are not like a battery running down but more like what happens to an overworked automobile engine (Moorcroft, 1993). Early on, compensation for the engine deficits can be made by gradually increasing pressure on the accelerator, just as a person can reduce the effects of sleep debt by extra effort and motivation, though eventually these methods of compensation will not be successful.

Other effects of sleep debt include experiencing negative emotions, forgetfulness, poor communication, apathy, decreased desire to socialize, lethargy, clumsiness, concentration difficulties, indecisiveness, and increased health problems (Maas, Axelrod, & Hogan, 1999). An in-depth discussion of these consequences will be presented in the sections, 'Individual consequences facet' and 'Organizational consequences facet'.

MEASURES OF SLEEPINESS

State of sleepiness can be measured by various objective and subjective methods. In this section, we first review two objective measures, polysomnography and pupillography, which record physiological activities related to the states of asleep and awake. After that, we review the most often used self-report measures, which can efficiently and effectively assess state of sleepiness.

Polysomnography records three physiological activities: brain waves (revealed by electroencephalogram or EEG), eye movements (revealed by electrooculogram or EOG), and neck muscle tension (revealed by electromyogram or EMG). The method works because many organs of the body generate small amounts of electrical energy as they perform their functions. Among these three types of electrical recordings, EEG provides the most noticeable distinctions among stages of sleep. The polysomnographic stages, as shown in Figure 3.2, are designated as awake (i.e., alert wakefulness and drowsy wakefulness), Stages 1–4 of sleep, and rapid-eye-movement sleep (REM sleep). The physiologies of Stages 1–4 are very similar and are often collectively referred to as NREM (i.e., non-REM). Furthermore, the distinction between Stages 3 and 4 is somewhat arbitrary, so they are collectively

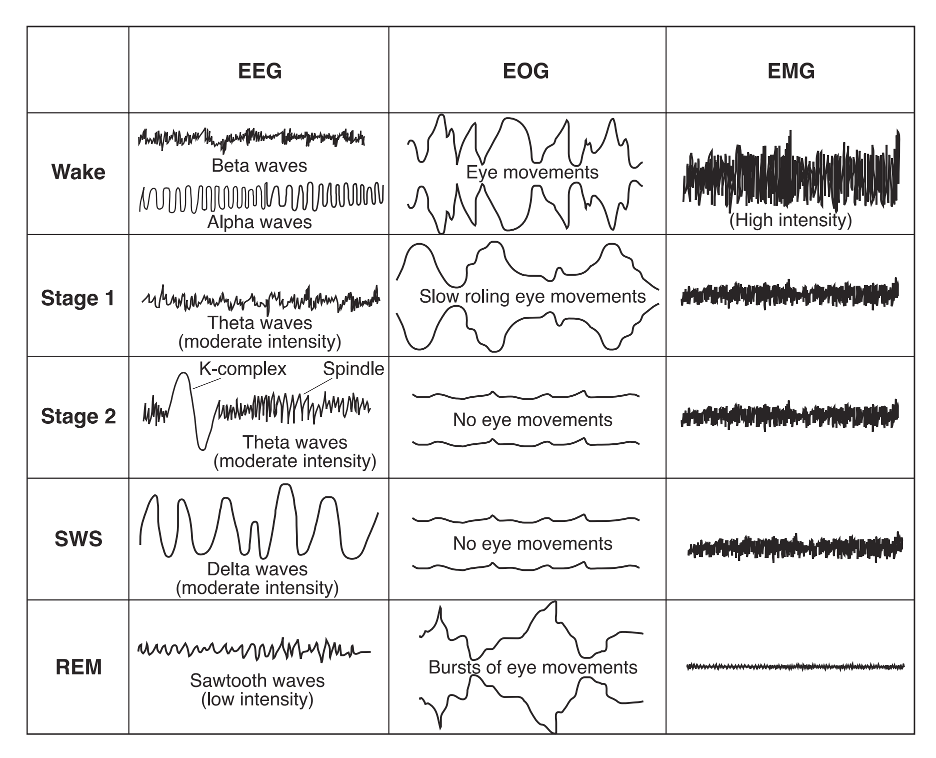


Figure 3.2 EEG, EOG, and EMG characteristics of waking and each stage of sleep. (*Note:* Beta waves = irregular low intensity, fast frequency (16–25 Hz) typically occurring in an awake, active brain. Alpha waves = regular moderate intensity, intermediate frequency (8–12 Hz) typically occurring in an awake but relaxed brain. Theta waves = moderate to low intensity, intermediate frequency (3–7 Hz). Delta waves = intense, low frequency ($\frac{1}{2}$ to 2–3 Hz), K-complex = large, slow peak followed by a smaller valley. Spindle = moderately intense, moderately fast (12–14 Hz) rythmic oscillation for $\frac{1}{2}$ to $1\frac{1}{2}$ seconds. Sawtooth waves = relatively low intensity, mixed frequency that often has a notched appearance. Waking eye movements tend to be relatively constant and have mainly sharp peaks and valleys with some smaller peaks and rounded peaks mixed in. Slow rolling eye movements are mostly large with rounded peaks. The eye movements of REM sleep usually have sharp peaks and come in bursts of a few seconds each with intervening quiet periods of a few to 10 seconds. The thickness of the EMG line is the key indicator.)

referred to as ‘slow wave sleep’ (SWS). REM sleep is a very unique stage of sleep. While the EEG during REM sleep closely resembles that of wakefulness, the muscles controlling body movements are paralyzed into a very relaxed state (as shown by the EMG). During REM sleep, the EOG shows bursts of rapid eye movements with seconds of quiescence between bursts.

As seen in Figure 3.2, awakening is easily identified by intense, high-frequency registrations on the EEG, EOG, and EMG recordings. In contrast, sleep onset is more difficult to identify because people do not

simply drop off to sleep. Instead, the transition from wakefulness to sleep is gradual and involves a complex succession of changes beginning with relaxed drowsiness, going through Stage 1, and ending in the first couple of minutes of Stage 2. NREM sleep represents the common understanding of sleep. The brain waves, especially during SWS, indicate both a relaxed brain and body idling together, however still capable of movement. Note that neither REM sleep nor NREM sleep is a quantitatively deeper sleep than the other. Rather, they are qualitatively different kinds of sleep. Instead of viewing the progress from NREM to REM sleep as a continuum moving away from awake, the two types of sleep should be viewed as different rooms in a house. Just as a kitchen differs from a family room that differs from a bedroom, so too does wakefulness differ from REM sleep that differs from NREM sleep.

An application of polysomnography is the examination of excessive daytime sleepiness using a series of naps at 2-hr intervals, referred to as the Multiple Sleep Latency Test (MSLT). Requiring usually 7 hours to complete, the MSLT consists of the following procedure: examinees are given a 20-min opportunity to fall asleep in a quiet, comfortable sleep lab every 2 hours during the day. They are instructed not to resist the onset of sleep. Between sleep times, they are free to read, write letters, watch television, or have visitors. Another physiological approach to measure sleepiness refers to pupillographic assessment. This method, conducted in darkness by means of an infrared-sensitive video camera, measures various indicators of sleepiness such as average pupil size, pupillary instability, and pupil diameter. For instance, pupils constrict and become unstable during sleep onset (Mitler, Carskadon, & Hirshkowitz, 2000). The MSLT and pupillographic assessment correspond highly and seem to reflect the same aspect of central nervous system activation (Danker-Hopfe et al., 2001).

In contrast to these physiological assessment tools, self-report measures take less time to evaluate sleepiness. The practical and economical advantages associated with self-report measures allow researchers and practitioners to screen people efficiently and prioritize patients for treatment (Pouliot, Peters, Neufeld, & Kryger, 1997). We will review the psychometric quality of three measures widely used to assess current state of sleepiness. The specific items of most of these scales can be obtained from Benca and Kwapil (2000) as well as Schutte and Malouff (1995).

Epworth Sleepiness Scale

The Epworth sleepiness scale (ESS), developed by Johns (1991), assesses the general level of daytime sleepiness or the average sleep propensity. It presents eight commonly encountered situations (e.g., sitting and resting, in a car while stopped for a few minutes in traffic), and respondents are asked to rate the likelihood that they would doze off or fall asleep on the basis of four response categories, varying from ‘would never doze’ (0) to ‘high chance of

dozing' (3). Test–retest reliability with five months apart in a normal sample was 0.82, and no mean difference on ESS scores was found (Johns, 1992). Johns also reported that the Cronbach alphas in two different samples ranged from 0.73 to 0.88. ESS scores have been related to obstructive sleep apnea and other objective indices of sleep problems including respiratory disturbance, oxygen saturation, polysomnography, and MSLT (Chervin, Aldrich, & Pickett, 1997; Johns, 1994; Pouliot et al., 1997).

Sleep/Wake Activity Inventory

The Sleep/Wake Activity Inventory (SWAI) developed by Rosenthal, Roehrs, and Roth (1993), consists of 59 items with 9 response categories, ranging from 'always' (1) to 'never' (9). Of the 59 items, only the 9-item Excessive Daytime Sleepiness subscale is relevant to assess the state of sleepiness. Rosenthal et al. reported the Cronbach alpha of the subscale as 0.89. They also provided validity evidence by demonstrating that the subscale scores inversely related to hours of sleep during the preceding week and positively related to the ease of falling asleep at night (Breslau, Roth, Rosenthal, & Andreski 1997).

Stanford Sleepiness Scale

The Stanford Sleepiness Scale (SSS), one of the oldest measures, assesses an individual's current level of sleepiness (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). Similar to the behavioral-anchored rating scale format, the SSS consists of one item with seven different levels of sleepiness, ranging from 'feeling active and vital; alert; wide-awake' to 'almost in a reverie; sleep onset soon; lost struggle to remain awake', and respondents select one of the seven anchors to describe their current state of alertness. The SSS has been well validated on average people for its intended purpose. Alternative form reliability, assessed by agreement, was reported to be 88% (Hoddes, Dement, & Zarcone, 1972). Hoddes et al. (1973) provided validity evidence that sleep deprivation was positively related to an increase in SSS scores; however, no norms are available for comparison purposes. The results of convergent validity for the SSS have been mixed (Danker-Hopfe et al., 2001; Johnson, Freeman, Spinweber, & Gomez, 1991).

Other one-item sleepiness scales widely used are the visual analogue scale (VAS) and the Karolinska Sleepiness Scale (KSS, Akerstedt & Gillberg, 1990). There are various versions of the VAS, which originated in educational research (Freyd, 1923) and have been subsequently developed to assess mood (Folstein & Luria, 1973). Generally, the VAS requires respondents to indicate how they feel by placing a mark on a line between 'most alert' at one end and 'most sleepy' at the other. The KSS consists of nine anchors, ranging from 'extremely alert' (1) to 'very sleepy, great effort to keep

awake, fighting sleep' (9), and the respondent selects which anchor is representative of his current state. Validity evidence for the VAS and KSS has been provided by Gillberg, Kecklund, and Akerstedt (1994). In general, reliabilities of one-item scales are difficult to estimate and tend to be low.

There are other self-report measures developed to evaluate sleep quality, which include the Pittsburgh Sleep Quality Index (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989), the St Mary's Hospital Sleep Questionnaire (Ellis et al., 1981), and the post-sleep inventory (Webb, Bonnet, & Blume, 1976). Because these scales do not assess current state of sleepiness, a review of their psychometric quality is beyond the scope of this chapter.

Subjective measures of sleepiness might not always be accurate, because people may not be aware of their true level of sleepiness due to a stimulating environment or high motivation that supersedes any potential experience of sleepiness. It is safe to conclude that the MSLT and other physiological measures such as pupillography assessment directly measure the pure physiological need to sleep, and the Maintenance of Wakefulness Test (MWT, Mitler, Gujavarty, & Browman, 1982) directly measures physiological attempts to remain awake. These measures eliminate psychological factors such as motivation that might prevent perceived sleepiness. On the other hand, the physiological measures may not show the extent to which sleepiness might occur in real-life situations. Now that the basics of daytime sleepiness and the common methods used to measure sleepiness have been discussed, we turn our attention to the antecedents and consequences of sleepiness in the workplace.

Because sleepiness in the workplace has not been systematically investigated, we applied the facet analysis approach to guide our following reviews and the development of a conceptual model to describe plausible antecedents and consequences of sleepiness on the job. A facet is a conceptual dimension underlying a set of mutually exclusive variables (Beehr & Newman, 1978). Beehr and Newman have suggested generating all possible facets and accompanying variables that are deemed relevant to the topic of interest, regardless of whether empirical evidence exists for the facets or specified variables.

The facet analysis resulted in four main facets: individual antecedents facet, organizational antecedents facet, individual consequences facet, and organizational consequences facet, as presented in Table 3.1. The individual antecedents facet includes demographics, health conditions, and personality, all of which are postulated to influence the extent to which individuals feel sleepy in the workplace. For instance, an employee's health condition may influence his sleep duration or sleep quality, which may in turn affect the level of alertness on the job. The organizational antecedents facet includes task characteristics, work schedule, job stressors, and physical environment, among other things. It is proposed that differences in these antecedents could lead to differences in the level of sleepiness experienced by individuals at work. For instance, people who experience a lot of stressors on the job may

Table 3.1 Facets of sleepiness in the workplace.

(1) Individual Antecedents Facet

- (a) Demographics
  - (i) Age
  - (ii) Gender
  - (iii) Ethnicity
  - (iv) Culture
- (b) Circadian rhythm
  - (i) Phase (morning–evening orientation)
  - (ii) Stability (rigidity/flexibility)
  - (iii) Amplitude (vigor/languidity)
- (c) Shorter or longer sleepers
- (d) Health conditions
  - (i) Body mass index
  - (ii) Minor illness
  - (iii) Pregnancy/Menopause
  - (iv) Psychiatric problems
  - (v) Other medical problems (e.g., sleep disorders, arthritis, osteoporosis, heartburn, gastroesophageal reflux, chronic obstructive pulmonary disease, congestive heart failure, collagen vascular disease, etc.)
- (e) Personality
  - (i) Locus of control
  - (ii) Neuroticism/Anxiety
  - (iii) Intraversion/Extraversion
  - (iv) Anger
  - (v) Depression
  - (vi) Conscientiousness
  - (vii) Type A/B
- (f) Work schedule experience
  - (i) 10- or 12-hr work schedule experience
  - (ii) Shift work experience
- (g) Family interfering with work (FIW)
  - (i) Number of dependents
  - (ii) Household work
  - (iii) Care tending

(2) Organizational Antecedents Facet

- (a) Task characteristics
  - (i) Prolonged vigilance
  - (ii) Task duration
  - (iii) Monotony
- (b) Work schedule
  - (i) Number of hours
  - (ii) Change in work schedule
  - (iii) Flextime
  - (iv) On-call work
  - (v) Shift work
    - Night work
    - Rotating shift

(d) Physical environment
  (i) Temperature
  (ii) Noise
  (iii) Lighting
  (iv) Humidity

(e) Work interfering with family (WIF)

(f) Management (leadership)

(g) Organizational/supervisory support

(h) Organizational values or climate (e.g., safety climate)

(i) Office design (desk, chair, windows, posters, etc.)

(j) Operational procedures

(k) Commuting distances

(3) Individual Consequences Facet

(a) Psychological aspects
  (i) Well-being
    - General affect/mood
    - Anxiety
    - Irritability
    - Depression
    - Feeling of alienation
  (ii) Satisfaction
    - Job satisfaction
    - Life satisfaction
  (iii) Motivation

(b) Physiological aspects
  (i) Sleep patterns
    - Sleep duration
    - Sleep quality
    - Circadian timing of sleep
    - Duration of prior wakefulness
  (ii) Subjective fatigue
  (iii) Physical symptoms

(c) Behavioral aspects
  (i) Coping ability
  (ii) Diet
  (iii) Drug use
  (iv) Smoking
  (v) Poor interpersonal relationship

(4) Organizational Consequences Facet

(a) Job performance
  (i) Task performance/productivity
  (ii) Absenteeism
  (iii) Accidents
  (iv) Injuries
  (v) Errors
(b) Creativity

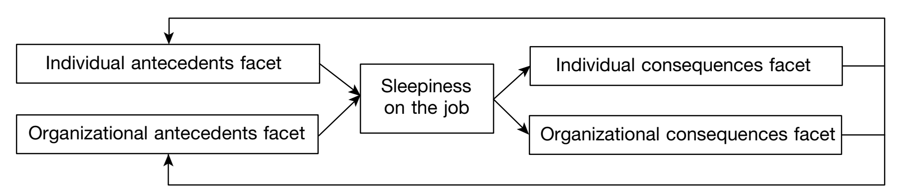


Figure 3.3 Conceptual model of antecedents and consequences of sleepiness in the workplace.

find it difficult to fall asleep at night, which could result in feeling sleepy during the day at work. The individual consequences facet ranges from psychological aspects to physical aspects to behavioral aspects. The organizational consequences facet includes many domains of job performance (e.g., task performance, accidents) and creativity.

The conceptual model depicting the relationships among these facets with sleepiness in the workplace is presented in Figure 3.3. In the model, we postulate that individual and organizational antecedents affect the state of sleepiness in the workplace, which results in both individual and organizational consequences. These consequences are further expected to recursively affect the individual and organizational antecedents.

INDIVIDUAL ANTECEDENTS FACET

According to the conceptual model depicted in Figure 3.3, individual characteristics such as personality and demographics may play important roles in predicting workplace sleepiness. In this section, we review pertinent literature with the foci on age, gender, ethnicity, personality, and health conditions.

Age

Age differences have been shown to exist in both the biological clock and sleep quota. The peak of sleepiness tends to shift later by an hour or two from adolescence to young adulthood and then generally shifts earlier by an hour or two with increasing age beyond that of young adulthood (Moorcroft, 2003). This is consistent with the fact that older individuals tend to be more morning-oriented (Akerstedt & Torsvall, 1981). By age 60 or 70, many adults experience a decrease in the proportion of time spent in the NREM sleep stage; however, the percentage of REM sleep remains relatively stable (Moorcroft, 2003).

The relationship between age and sleepiness on the job has not been systematically studied, with a few exceptions. Lee (1992) documented a positive relationship between age and sleep disturbances. It has been found that subjective sleepiness during night shift work increases with age (Seo,

Matsumoto, Park, Shinkoda, & Noh, 2000). Akerstedt and Torsvall (1981) and Parkes (1994) also showed that both sleep quantity and quality decreased with increasing age and experience with shift work. Parkes further revealed that older workers experienced greater difficulty in adjusting to shift work than younger workers.

An additional finding from Parkes (1994) was that the relationship between age and quantity of sleep was stronger than that between age and quality of sleep. While the trend of decreased sleep quality with age has been seen in other research (Marquie, Foret, & Queinnec, 1999), the decrease in sleep quantity is a much more common result. Though sleep duration after night work decreased with age, older workers did not report more sleep difficulties (Seo et al., 2000; Spelten, Totterdell, Barton, & Folkard, 1995). Furthermore, the shorter sleep duration did not have an effect on older workers' overall on-shift alertness. In fact, the older workers had higher on-shift alertness than the younger workers. At first glance, this result seems to be inconsistent with Parkes' results. However, this discrepancy may be attributed to the fact that Parkes was investigating the adaptation to shift work by older and younger workers, and Spelten et al. were studying people whose shift work tenure was longer. It may be more difficult for older workers to adjust to shift work than younger workers consistent with the findings by Parkes. Those older workers studied by Spelten et al. have been engaging in shift work for a substantial period of time and have probably adequately adapted since they still remain on this schedule. This idea is consistent with the concept of the 'healthy worker' effect and is confirmed by Bourdouxhe et al. (1999). Considering both current and former employees, Bourdouxhe et al. found that increased age of current workers was not related to increased sleep problems; however, age and sleep problems of former employees were related. It seems that aging employees who did have sleep problems had left shift work.

In a study of the effects of sleep deprivation on recovery sleep, Gaudreau, Morettini, Lavoie, and Carrier (2001) reported that middle-aged participants showed a decrease in their ability to maintain sleep during an abnormal circadian phase (sleep during the day instead of at night) when compared with younger people. This problem can be attributed to a reduction in homeostatic recuperative drive while aging, which might explain increases in complaints related to shift work among middle-aged people.

Gender

Based on the Women and Sleep Poll (NSF, 1998), the average woman aged 30–60 sleeps about six hours and forty-one minutes during the workweek and possesses better sleep–wake patterns than men (Jean-Louis, Kripke, Assmus, & Langer, 2000). Conditions unique to women such as the menstrual cycle, pregnancy, and menopause (changes of estrogen and progesterone), can affect

how well a woman sleeps. In fact, the poll further found that 50% of menstruating women reported disrupted sleep for two to three days each cycle. If this condition and the others mentioned above are causes of sleep debt, a consequence may be sleepiness on the job.

Within limited empirical studies, there is no strong evidence suggesting a gender difference in workplace sleepiness. Caldwell and LeDuc (1998) did not find any significant differences in flight performance or recovery sleep between female and male sleep-deprived pilots, but did observe that the males were generally more anxious than the females. No significant gender difference in performance during shift work has also been found, suggesting that both men and women may be equally capable of adjusting their sleep to accommodate that type of work schedule (Beerman & Nachreiner, 1995).

Ethnicity

There has been little if any research done on cultural or ethnic differences in sleep patterns as they relate to work. Jean-Louis, Kripke, and Ancoli-Israel (2000) did examine ethnic differences in sleep patterns along with the gender differences described above and found that men of minority races reported the worst sleep quality.

Personality

Similar to the other individual difference characteristics, research exploring the relationships between personality variables and sleepiness in the workplace has been relatively limited. Most of the work reviewed below concentrates on how personality is related to sleep quantity, sleep quality, and sleep disturbances rather than actual sleepiness on the job.

One relationship that has received a fair amount of attention is that between neuroticism and sleep duration. Most researchers have hypothesized that more neurotic individuals are most likely to sleep less. For the most part, findings have been consistent with this contention (Kumar & Vaidya, 1982). A recent study conducted by Gray and Watson (2002) investigated the relationships between all of the Big Five personality characteristics (i.e., neuroticism, extraversion, conscientiousness, agreeableness, openness to experience) and three components of sleep (i.e., sleep quantity, quality, schedule). While they found no significant relationships between personality and quantity of sleep, they did find that sleep quality was negatively related to neuroticism and positively related to both extraversion and conscientiousness. Parkes (1999) also reported a similar finding that neuroticism positively related to sleep problems. In the area of sleep schedule, the strongest correlate was conscientiousness. Specifically, those individuals scoring lower on conscientiousness had a tendency to maintain 'evening-oriented' schedules.

Personality and sleep may also be related such that the severity of consequences associated with sleep loss may differ dependent upon personality. Though minimal evidence is available to support this proposition, Blagrove and Akehurst (2001) did find that mood deficits associated with sleep loss were more severe for those high in neuroticism or extraversion.

Consistent with Kumar and Vaidya (1982), Type A individuals have reported more sleep problems (Parkes, 1999), more trouble falling asleep, more nightmares, and less time asleep compared with Type B individuals (Koulack & Nesca, 1992). Finally, the relationship between locus of control and sleep has been touched upon briefly in the literature, though the findings are contradictory. Although Hicks and Pellengrini (1978) found that long-sleepers viewed themselves as more internally regulated than short-sleepers, Kumar and Vaidya (1986) found an association between external locus of control and longer sleep duration.

Morning-Evening Orientation

Morning–evening orientation, or 'morningness', refers to whether individuals prefer being more active in the morning or evening and represents one type of the circadian rhythm. In general, most people are intermediate types and have no extremely strong preference for either morning or evening activity. Most of the research that considers morningness as an individual characteristic has investigated the tolerance of people more or less morning-oriented for different work schedules.

Seo et al. (2000) found that when working the day shift, morning types tended to go to sleep earlier and wake earlier than evening and intermediate types. Findings concerning the night shift were also as expected. Specifically, a greater percentage of morning types reported feeling sleepy earlier and higher general levels of sleepiness during the night shift when compared with evening and intermediate types. Khaleque (1999) further demonstrated that workers characterized as evening types reported better sleep quality regardless of the shift they worked compared with morning types. These findings suggest that evening types may be better suited for not simply night work but rather shift work in general.

While most individual differences are generally conceptualized as stable, there is some evidence that morning–evening orientation may be somewhat malleable. Mecacci and Zani (1983) found that adult workers tended to be more morning-oriented than college students. They suspected that people might possess the capability to alter their sleep–wake pattern when a reason presents itself such as starting a job. Indeed, some evidence exists that people can adjust their morning–evening orientation to some extent, except for those who possess extreme orientations (Hildebrandt & Stratmann, 1979).

Work Schedule Experience

Evidence exists that experience with shift work might mitigate the negative effects of the work schedule on sleep. Shift work experience has been positively related to adjustment of sleeping pattern and sleep quantity as well as negatively related to perceptions of the shift being strenuous and tiresome (Breaugh, 1983; Parkes, 1994).

Family Interfering with Work (FIW)

Individuals with significant familial commitments most likely experience considerable sleep debt and, in turn, sleepiness at work. Indeed, dual-working couples who are also caregivers report averaging slightly less sleep than do those who are non-caregivers (NSF, 1998–2002). Spelten et al. (1995) found that, as the number of dependents in the household and the level of perceived work–home conflict increased, so did sleep difficulties. These same participants reported decreased sleep duration and alertness on the job. A differential relationship may exist between FIW and sleepiness at work among men and women, considering that women oftentimes still bear the brunt of the workload at home.

Health Conditions

Individual physical health may affect sleep patterns, which in turn might influence sleepiness at work (Haermae, 1993). For instance, minor illness or pregnancy could temporarily increase the body's sleep quota. Certain medical problems (e.g., arthritis, heartburn, osteoporosis, heart disease) as well as psychological conditions such as depression and anxiety may affect sleep adversely. For example, individuals suffering from arthritis may have difficulty falling asleep or may be awakened by painful joints. Furthermore, the occurrence of heartburn and regurgitation during sleep as a result of Gastroesophageal Reflux (GER) may cause sleep debt, which later could lead to sleepiness on the job. It should be noted that many medical problems are more common in older people; therefore, interactive effects of age and health on workplace sleepiness are likely.

An important point to consider is that the variables reviewed earlier may serve other functions besides being antecedents of workplace sleepiness in the simplest form. For instance, age may moderate the relationship between morning–evening orientation and workplace sleepiness such that older individuals are more likely to be morning-oriented and subsequently experience more sleepiness when engaging in night work. Besides individual characteristics interacting with each other in their effects on sleepiness, other variables pertinent to the organization may interact with individual characteristics as

well as affect workplace sleepiness on their own. It is the organizational antecedents that we turn to next.

ORGANIZATIONAL ANTECEDENTS FACET

Myriad variables associated with the organization can be posited to affect sleepiness in the workplace. Since no systematic consideration of these potential variables has been previously completed, the research to support the link between the organizational variable and sleepiness may be considerable in some cases and sparse in others. The following reviews provide evidence that organizational factors have the potential to play a large role in affecting workplace sleepiness.

Task Characteristics

Some job tasks may cause the employee to experience sleepiness while they perform them. As a result, sleepiness in the workplace is most likely to coincide with certain occupations that perform these tasks. Ironically, the occupations performing the tasks likely to induce sleepiness are also the ones in which performance decrement in the form of an error could be extremely costly. Types of tasks that have been associated with sleepiness are those requiring monotonous movements, a long duration of time, or prolonged vigilance. Occupations in which these types of tasks are commonly performed include air traffic controller, professional driver, pilot, policeman, security/prison guard, and combatant (for an example of how sleep is a factor for a pilot, see Nicholson, 1987; for a coach driver, see Sluiter, van der Beek, & Frings-Dresen, 1999).

A large-scale study was conducted on flight crews in which their sleep, circadian rhythms, subjective fatigue, mood, nutrition, and physical symptoms were monitored before, during, and after flight operations (Gander, Rosekind, & Gregory, 1998). Findings showed that, while all of these variables were affected during the operations to some extent, the type of operation (e.g., short-haul fixed-wing vs. long-haul) played a large role in the magnitude of the effects with most detrimental effects occurring on those requiring the longest amounts of time.

Work Schedule

An employee's work schedule is an obvious variable that may affect sleepiness in the workplace. The specific features of a work schedule considered here are the number of hours worked, change in schedule, on-call work, and shift work.

Number of hours

Working long hours may be associated with sleepiness while at work. Number of hours worked has been related to reports of fatigue (Lilley, Feyer, Kirk, & Gander, 2002), and people working a compressed schedule (i.e., 10 hours a day or longer) have reported poorer subjective health, well-being, and sleep quality compared with workers on an 8-hr day shift (Martens, Nijhuis, van Boxtel, & Knottnerus, 1999). In a recent meta-analysis, Sparks, Cooper, Fried, and Shirom (1997) found a small, but significant, positive mean correlation between work hours and overall adverse health. When distinguishing between physiological and psychological health symptoms, the authors found that the mean correlation was larger between work hours and psychological health symptoms (e.g., poor sleep) than that with physiological health symptoms.

Studies examining the effects of increasing the workday from 8 hours to 12 hours have obtained conflicting results. One study found that the 4-hr increase to the workday resulted in workers experiencing considerable sleep debt and subsequent decreased performance and alertness (Rosa, 1991). A follow-up study conducted 3 to 5 years after this initial research revealed that these original detrimental effects of the 12-hr workday persisted. Other research has concurred regarding the unfavorable results of this longer workday, suggesting that sleepiness is greater during a 12-hr shift than an 8-hr shift and is accompanied by feelings of fatigue and decreased alertness. As these states tend to accumulate with the progression of the workweek, the result is often errors in work and judgment.

In contrast to the above negative findings, Lowden, Kecklund, Axelsson, and Akerstedt (1998) concluded that workers responded favorably to the change in the forms of decreased sleepiness while at work and increased sleep, recovery time after night work, and satisfaction with work hours. No significant difference in job performance was found between the two workday lengths. Other employees have also reported salubrious effects on sleep and psychological variables when their workday was changed from 8 hours to 12 hours (Mitchell & Williamson, 2000). Specifically, the increase in workday resulted in improvements in the following areas: sleep duration, uninterrupted sleep, sleep quality, mood, work schedule satisfaction, physical health symptoms (e.g., headaches), use of sleep aides, and social and domestic life satisfaction. No differences were observed in cognitive performance, though errors in dealing with unexpected situations increased during the final hours of the 12-hr shift.

The contradictory findings of these studies suggest that other variables might be interacting to determine whether employees react favorably or unfavorably to the changes of workday length. Though no conclusion can be made regarding the effects of a longer workday, strong implications of this research exist given the increases in both long work hours in the general

business world and flextime schedules involving long hours over a shortened workweek.

Change in schedule

Research has shown that only small changes in schedule may produce considerable alterations in a person's sleep quality, affect, and performance. Monk and Aplin (1980) used the natural changes during spring and autumn daylight saving times to investigate this phenomenon. It seems that adjustments to both the spring and autumn time changes required around a week to occur. In addition, while the spring adjustment period was associated with negative mood, the fall adjustment was associated with positive mood, increased perceptions of sleep quality, and even increased performance on a cognitive task during the morning hours. The above findings suggest that changing a shift start time to one hour earlier or later may hamper or facilitate behavior, respectively, although these effects as a result of the time change may only be temporary.

On-call work

'On-call' work consists of an employee being available to be called in to work at any time during a designated period. A profession often required to work a significant amount of time on-call is that of doctors. By using a longitudinal design, Lingenfelser et al. (1994) have examined the effects of on-call work on a host of variables. Doctors experienced decreased neuropsychological and cognitive functioning as well as more negative mood after being on-call for 24 hours compared with after a night off duty and after a period of uninterrupted sleep. On-call work is also common in the railroad industry, such that engineers are often called in to operate a train when a shortage occurs. Pilcher and Coplen (2000) found that although on-call engineers showed no difference in sleep quantity compared with those working regular schedules, they reported worse sleep quality in the forms of difficulty going to sleep and inability to stay asleep. It seems that on-call work may have deleterious effects on employees' sleep as well as other variables, but more research needs to be done in order to understand the full effects.

Shift work

A shift-worker is someone who works at a time inconsistent with the natural circadian rhythm (i.e., any shift other than the common day shift). A voluminous amount of research has been conducted in the area of shift work dating back to the 1950s when the manufacturing industry first initiated continuous production. The basic finding of this early research was that the most fundamental effects of shift work were on the worker's sleep quantity

and sleep quality (Hurrell & Colligan, 1986). There is concern about the myriad detrimental effects associated with shift work, especially given the increase in continuous shifts to boost productivity and the creation of flex-time schedules to accommodate employees' familial obligations (Kogi, 1991). Dawson and Fletcher (2001) confirmed that all types of shift work schedule resulted in significantly higher amounts of work-related fatigue compared with the standard work schedule. In general, shift-workers experience significant decreases in mood, health, mental skills, and performance and higher incidence of sleep disorders, emotional problems, stomach and intestinal problems, and cardiovascular illnesses. They also have a higher than average number of car accidents when driving home from work (Harrington, 1994).

Shift work may be undesirable not only because of the physiological consequences but also because of the social consequences associated with working at irregular times while others are engaging in social, religious, recreational, and entertainment activities. Frost and Jamal (1979) described this concept as low compatibility between work and non-work and found that it was associated with low levels of need fulfillment at work, social involvement, and emotional well-being as well as high levels of anticipated turnover. Shift-workers often complain of social isolation and have 57% higher divorce rates than non-shift workers (Moorcroft, 2003).

Social and domestic pressures to be active rather than sleeping during their time off may spur shift-workers to participate in activities during the day even if they are sleep-deprived and synchronized to a different schedule, resulting in further sleep complications and sleepiness while at work. This may be especially relevant for women shift-workers who are expected to run the household on a 'normal' schedule, but it can also affect men as they attempt to fulfill their roles as sex partner, social companion, and father.

Akerstedt (1990) pointed out that different schedules of shift work might have differential effects on sleep, because some shifts may be more in line with normal sleeping patterns than others. For instance, individuals working the night shift are on the job during the lowest alertness point of their circadian rhythm. Though they fall asleep rapidly after their shift is over, they are awakened too early because of their circadian rhythm, and the effects spill over to the following night shift. In the case of early morning workers, their circadian rhythm makes it difficult to fall asleep early the preceding night, which affects them during their morning shift. This inconsistency between circadian rhythmicity and the sleep–wake cycle for night and morning workers can result in extreme sleepiness while on the job. Even more negative consequences may be experienced by those who work rotating shifts, because of the continual disruption to the circadian rhythm and sleep–wake cycle. Because of the severity of outcomes associated with night shift and rotating shift work, empirical findings related specifically to these two work schedules are reviewed in detail below.

Night shift. Deleterious effects of night shift on various psychological and physical indicators have been reported in multiple studies. Breaugh (1983) found that those who worked a shift from noon to midnight reported less sleep problems (both quantity and quality) than those who worked a shift from midnight to noon. The negative effects of night shift include the following: increased sleepiness, reduced alertness, worsened mood, impaired performance in the forms of slower speed and less accuracy, and increased risk of fatigue-related mistakes, accidents, and injuries (Akerstedt, 1995; Bohle & Tilley, 1993; Folkard & Monk, 1979). Long-standing evidence has existed that performance is worse at night compared with during the day (Colquhoun, 1971). Once the sleep–wake cycle is disrupted, a sharp decline in work efficiency is usually observed, with this deficit having the tendency to level off after approximately one week. One laboratory study simulated night work by allowing the participants to sleep during the day. Findings showed that performance on simple visual-acuity tasks at night was not affected, while performance on cognitive and monotonous tasks requiring a high level of attention and long duration of time was considerably degraded and associated with sleepiness (Porcu, Bellatreccia, Ferrara, & Casagrande, 1998).

Furthermore, one study found that night shift-workers reported more subjective health complaints than those working the common day shift (Martens et al., 1999). Folkard and Monk (1979) also discuss the potential for situational constraints to interact with work schedules on performance such that those working at night may experience a lack of resources or be forced to use poorer equipment.

Rotating shift. Rotating shifts may be the most perilous to workers in terms of both adverse psychological and physical consequences. The most commonly reported disadvantages associated with a work schedule consisting of 12-hr shifts rotating from day to night were chronic fatigue, impaired physical recovery after the shift, and sleep disorders (Bourdouxhe et al., 1999). In comparison with day-workers and those working permanent shifts, a rotating shift work schedule tends to be associated with decreased general health, well-being, quality of sleep; disruption of sleep, family life, social life, leisure activities, regularity of meal time, and digestive system functions (Czeisler, Moore-Ede, & Coleman, 1982; Khaleque, 1999; Martens et al., 1999).

It is important to note here that the health impairment of these rotating shift-workers may not be solely a consequence of sleep debt but also a combined outcome of the multiple psychosocial factors adversely impacted by the work schedule. Those working a rotating shift schedule were more likely to fall asleep on the job, feel fatigued, and experience confusion along with decreased levels of vigor and activity during the night phase compared with the other phases of the schedule (Luna, French, & Mitcha, 1997). The mood and reaction times of rotating shift-workers also decreased over the

course of the night shift (Totterdell, Spelton, Barton, Smith, & Folkard, 1995).

Job Stressors

A common topic investigated in the stress literature is the effect of a stressor on well-being, in which sleepiness is a corollary. Some studies have looked at the relationship between stressors and particularly sleep-oriented variables. Stressors have been shown to be associated with depth of sleep, difficulties in waking up, quality and latency of sleep, and sleep irregularity (Verlander, Benedict, & Hanson, 1999). Van Reeth et al. (2000) concluded that both acute and chronic stressors have pronounced effects on sleep architecture and circadian rhythms. Increased perceptions of general work stress have been associated with insomnia and job-related burnout, which is characterized by sleep disturbance (Hillhouse, Adler, & Walters, 2000). Besides stressors presumably having direct effects on sleep, the reactions to stressors may have negative consequences on sleep patterns. Specifically, if psychological or physiological reactions toward stressors are prolonged and uncontrollable, they may cause abnormal hypothalamo-pituitary-adrenal secretory activity, which results in ineffective regulation of the sleep–wake cycle. Research supporting the relationships between particular types of job stressors and sleep is presented below.

Job strain model

Job demand and job control (or job discretion) are two stressors specifically related to work that receive a lot of attention. The combination of high job demand and low job control results in job strain, which is characterized by poor psychological and physical well-being (Karasek, 1979). Parkes (1999) reported small but significant positive relationships between job demand and lack of job control with sleep problems. Generally, research linking job demand to sleep is sparse, and studies that do investigate this relationship usually focus on one job characteristic that creates a high level of demand (e.g., time pressure). For instance, the intensive pace of work has been associated with high levels of fatigue (Lilley et al., 2002), workload has been negatively related to sleep quality (Martens et al., 1999) and somatic symptoms (e.g., trouble sleeping, Spector & Jex, 1998), and time pressure at work has been positively related to sleeping pill consumption for females (Jacquinet-Salord, Lang, Fouriaud, Nicoulet, & Bingham, 1993).

Job demand, measured by the number of hours at the wheel for coach drivers, predicted quantity and quality of sleep as well as frequencies of stimulant consumption at work and alcohol consumption at night in order to stay awake and fall asleep, respectively (Raggatt, 1991). Job demand was also associated with difficulty falling asleep, difficulty staying asleep, difficulty getting back to sleep, and unintentional early morning awakening

(Marquie et al., 1999). Comparing Finnish and US managers on work perceptions and health symptoms, Lindstroem and Hurrell (1992) showed that US managers experienced higher levels of job demand than Finnish managers. In addition, US managers reported greater sleep problems than Finnish managers. These results suggest that the prevalence of both job demand and sleep problems may be nation-specific.

The combination of high job demand and low job control has been related to insomnia, sleep deprivation, daytime fatigue, psychosomatic and health complaints, low well-being, poor sleep quality, and emotional exhaustion (Kalimo, Tenkanen, Haermae, Poppius, & Heinsalmi, 2000; Sluiter et al., 1999). There is evidence that these relationships may remain stable regardless of the number of hours worked by the employee or his/her lifestyle. Though the sleep constructs considered thus far have been mainly subjective in nature, some support exists that high demand and low control may be related to the physiological characteristics of sleep, particularly to the increase of systolic blood pressure during sleep (van Egeren, 1992). Contrary to this finding, Rau, Georgiades, Fredrikson, Lemne, & de Faire (2001) observed no effect of high demand and low control on heart rate or blood pressure during sleep, though they did find that lower perceptions of job control alone were associated with increased heart rate and diastolic blood pressure at night.

Interpersonal conflict

Interpersonal conflict at work may be a plausible cause of a sleepless night. A study conducted by Bergmann and Volkema (1994) examined the most common work conflict issues, behavioral responses to the conflicts, and consequences of the conflicts. The second most common consequence of an interpersonal work conflict was 'lost sleep'. This consequence was experienced most often when the other party in the conflict possessed legitimate power (e.g., a supervisor) and either emotional or withdrawal behaviors were involved in the conflict (e.g., crying, resigning). Spector and Jex (1998) also reported a positive mean correlation between interpersonal conflict and somatic symptoms in a meta-analysis. Similarly, Vartia (2001) found that the targets of workplace bullying as well as bystanders experienced greater mental stress (e.g., staying awake at night) than those not involved in bullying. Furthermore, victims of workplace bullying reported taking more sleep-inducing drugs and sedatives than both observers of bullying and non-bullied employees.

In addition to the above chronic job stressors, others may also be associated with poor sleep at night and subsequent sleepiness in the workplace. Spector and Jex (1998) reported a positive mean correlation between somatic symptoms and organizational constraints (i.e., situations that prevent employees from accomplishing their tasks). Perceived safety in police work

was also significantly related to poor sleep quality (Neylan, Metzler, Best, Weiss, Fagan, et al., 2002). Furthermore, poor atmosphere at work related to experiencing sleep disturbances and using sleeping pills to aid onset of sleep (Jacquinet-Salord et al., 1993).

Acute stressors

A consequence of an acute stressful experience (e.g., post-shooting, layoffs, disasters) may be disturbances in sleep patterns, which in turn could result in workplace sleepiness (Farnill & Robertson, 1990). Raggatt (1991) found that the occurrence of acute events was related to the frequencies of taking pills to stay awake at work and drinking alcohol to fall asleep at night. More common in some professions than others, accidents, injuries, or other workplace incidents may serve as job stressors that affect sleep patterns. The experience of stressful work events by police officers (e.g., pursuit of an armed suspect) was associated with the occurrence of psychosomatic symptoms and negative states including insomnia (Burke, 1994). The strain experienced after air traffic incidents and the subsequent effects of this distress were assessed in a population of civil aviation pilots (Loewenthal et al., 2000). Findings demonstrated that air traffic incidents induced strain, which subsequently resulted in distress-induced sleep disturbances. These sleep disturbances were also shown to impair performance.

Physical Environment

Certain characteristics of the physical environment may have an effect on sleepiness. For example, Marquie et al. (1999) found that exposure to noise as well as exposure to heat, cold, and bad weather positively predicted difficulty falling asleep, difficulty staying asleep, difficulty getting back to sleep, and unintentional early morning awakening. These findings remained after controlling for age and work schedule (i.e., daytime worker or rotating shift-worker). Another study found that exposure to noise in the workplace had no significant relationship with subjective sleep disturbances or consumption of sleeping pills, though the authors suggested that these non-findings might be the result of noise levels not being an issue in the organizations in which the data were collected (Jacquinet-Salord et al., 1993).

Work Interfering with Family (WIF)

Work interfering with family (WIF) is an organizational antecedent that is expected to influence both psychological and physical consequences, though the research specifically examining sleep constructs is sparse. WIF has been shown to co-vary with negative states such as insomnia (Burke, 1994) and emotional exhaustion (Boles, Johnston, & Hair, 1997). Senecal, Vallerand, and Guay (2001) found strong predictability of work–family conflict for

emotional exhaustion measured such as, 'I felt exhausted when I came back to work' (p. 181). Other studies have found a similar result with regard to work–family conflict and burnout (Bacharach, Bamberger, & Conley, 1991). Burnout was measured with items that have predominant emphasis on sleep-oriented factors (e.g., being tired, being physically exhausted, periods of fatigue when you couldn't 'get going').

As illustrated in Figure 3.3, sleepiness on the job is expected to influence two broad outcomes, which have important financial, health, and legal implications for employees and organizations. We will first focus on possible consequences pertaining to employees, followed by those related to organizations. Once again, readers should bear in mind that there has been little research specifically investigating the effects of workplace sleepiness. Although most of the reviews presented below rely on studies that have examined the effects of sleep deprivation or fatigue, these findings should be appropriate to infer the consequences of sleepiness in the workplace.

INDIVIDUAL CONSEQUENCES FACET

The individual consequences facet can be distinguished into three components; namely, psychological aspects, physical aspects, and behavioral aspects. Psychological aspects represent variables such as negative affect and motivation; the primary physical consequence considered is health problems; and behavioral aspects reviewed include alcohol/drug consumption.

Psychological Aspects

The effects of sleepiness on psychological well-being is the topic of a considerable amount of research, primarily focusing on the effects of work schedules on well-being and affect; however, it is believed that work schedule is usually found to affect these psychological constructs indirectly through sleepiness (Scott & LaDou, 1990). For instance, Barton et al. (1995) found that the negative effect of the number of consecutive nights worked on psychological well-being was mediated by sleep duration and sleep quality. The stability of relationships between sleep quality and psychological aspects has been demonstrated over a three-month period (Pilcher & Ott, 1998). Findings of Jean-Louis, Kripke, and Ancoli-Israel (2000) support the notion that psychological well-being may be more related to sleep quality than sleep duration.

The consequences of sleep deprivation may include inability to control negative mood, excessive euphoria, immature or inappropriate behaviors, emotional outbursts as well as inability to display empathy (e.g., Berry & Webb, 1985; Kramer, Roehrs, & Roth, 1976). Empirical findings have shown that lack of sleep (Blagrove & Akehurst, 2001; Bugge, Opstad, & Magnus,

1979; Krueger, 1989; Totterdell, Reynolds, Parkinson, & Briner, 1994) and night shifts (Bohle & Tilley, 1993; Luna et al., 1997) are associated with negative mood. Job-related burnout characterized by sleep disruption has also been associated with mood disturbances (Hillhouse et al., 2000).

In addition to the effect on general psychological well-being or affect, sleepiness is also found to be negatively related to achievement motivation and a sense of coherence, and positively related to irritability (Dalbokova, Tzenova, & Ognjanova, 1995). Quality of sleep was also significantly related to job satisfaction (Jacquinet-Salord et al., 1993), frustration, anxiety, and intention to quit (Spector & Jex, 1998). Similarly, Raggatt (1991) reported that quantity and quality of sleep were positively related to job satisfaction and negatively related to psychological symptoms (e.g., depression).

Although some studies reviewed above utilized experimental or longitudinal designs, their findings of the association between sleep deprivation and psychological constructs (e.g., affect, well-being) should not be used to infer direct causal relationships between the variables. For example, sleep deprivation may co-vary with hormone changes, which may actually influence affect. Totterdell et al. (1994) substantiated that affect may influence subsequent sleep quality and sleep duration, although the causal relationships can also be recursive. Van Reeth et al. (2000) suggested that employees who suffer from chronic sleep deprivation experience distress about their jobs and lives, which in turn can have effects on later sleep.

Physical health aspects

Effects of sleep deprivation on general physical health from both animal and human research have been well documented (e.g., Appels & Schouten, 1991; Landis, Bergmann, Ismail, & Rechtschaffen, 1992). Similar to the findings regarding psychological well-being, sleep quality most likely has a stronger relationship with physical health than sleep quantity (Pilcher, Ginter, & Sadowsky, 1997). Indeed, sleep quality has been correlated with physical health symptoms such as digestive problems (Pilcher & Ott, 1998). Deleterious consequences of long-term sleep deprivation in rats included skin lesions on hairless regions, decreases in body weight despite dramatic increase in eating, deficient defense against infection, and deficits in body temperature regulation accompanied by excess heat loss (Rechtschaffen & Bergmann, 2002). Note that one of the functions of sleep is to reduce body temperature, and continuous periods of high body temperature can be detrimental.

Studies examining the relationship between sleep and well-being tend to consider both indices of psychological and physical well-being; as a result, some studies reviewed here may include results concerning both aspects. Workers with a 12-hr rotating work schedule experienced pronounced health symptoms such as digestive, cardiovascular, and psychological disorders (Bourdouxhe et al., 1999). These health problems are common to

those who work long and irregular shifts; in fact, the occurrence of these symptoms has been aptly labeled 'shift-worker syndrome'. Compared with those possessing good sleep patterns, poor sleep patterns have been associated with higher rates of health care utilization for lumber mill workers (Donaldson, Sussman, Dent, Severson, & Stoddard, 1999), and greater hospitalizations during tour of duty for Navy sailors (Johnson & Spinweber, 1982). Hillhouse et al. (2000) further substantiated that medical residents experiencing sleep disturbances also reported poorer levels of general health. It appears that the association of poor sleep with poor health condition spans occupational boundaries. Finally, Hoogendoorn et al. (2001) identified an association between sleep difficulties and low-back pain. Note that, while sleep difficulty was assessed prior to the onset of back pain in their study, it is also likely that sleep problems and back pain can reciprocally influence each other.

Behavioral aspects

Sleep disorders have been linked with violence and aggression (e.g., see the first case study reported by Guilleminault & Poyares, 2001); however, little research could be identified that specifically explored the relationship between sleep debt and interpersonal relationships. In a 14-day longitudinal study, an earlier onset of sleep predicted better social interaction experience (i.e., spending time with people) during the following day (Totterdell et al., 1994). Indirect evidence reported by Harrison and Horne (1997) suggested that sleep-deprived people might experience significant deterioration in word generation and in the use of appropriate voice intonation, which results in a more monotonic or flattened voice. These types of behavior may have important implications for interpersonal communication in the workplace.

Sleep has been found to be related to smoking behavior, although the direction of the relationship is not consistent. Some observed a positive relationship between smoking and sleep problems (e.g., Patten, Choi, Gillin, & Pierce, 2000), negative relationships between smoking behaviors and daytime sleepiness as well as sleep problems (e.g., Haermae, Tenkanen, Sjoeblom, Alikoski, & Heinsalmi, 1998), or no relationship between smoking habits and sleep problems based on unreported data (K. R. Parkes, personal communication, June 17, 2002). These inconsistent results suggest the need to examine potential moderators. As demonstrated by Parkes (2002) the relationship between sleep problems (e.g., duration) and smoking behavior varied contingent upon work schedule (e.g., day or night shift) as well as work environment (i.e., offshore or onshore) for oil/gas industrial workers. Specifically, sleep duration for day shift-workers was significantly shorter for smokers than non-smokers while they were working onshore.

In addition to smoking behavior, Raggatt (1991) pointed out that people who experience job stressors tend to take pills to stay awake at work and

drink alcohol to fall asleep at night. Drinking alcohol, viewed as a passive coping behavior, was also found to be related to psychosomatic symptoms (e.g., sleep problems: Lindstroem & Hurrell, 1992). Furthermore, Haermae et al. (1998) reported that the relationship between alcohol consumption and sleep complaints was stronger for employees who work second shift, third shift, or an irregular shift schedule.

ORGANIZATIONAL CONSEQUENCES FACET

As shown in Table 3.1, the organizational consequences facet consists of both job performance and creativity. We review the relationships between sleepiness and major job performance criteria as outlined by Smith (1976). The specific criteria examined are task performance, absenteeism, and safety (i.e., errors, injuries, and accidents). A short discussion of creativity follows the review of job performance.

Job Performance

Task performance

An NSF poll conducted in 2000 estimated that workplace sleepiness costs US employers about $18 billion per year due to lost productivity. Indeed, sleepiness on the job has been associated with difficulty in concentration and inefficiency when solving problems and making decisions (Alapin et al., 2000). Furthermore, workplace sleepiness may be related to difficulty in handling stress, which has a significant implication for jobs in which ambiguity, time pressure, or emergencies are common. A combination of sleepiness and stress may compound the level of performance impairment. Increased distraction and reduced alertness have been reported by workers who were experiencing high levels of both on-the-job sleepiness and work stress (Dalbokova et al., 1995).

The type of task may interact with sleep debt such that engaging in particular kinds of tasks when sleepy may result in greater performance decrements. Specifically, sleep debt may result in delayed responding or complete failure to respond when engaging in tasks that are long, monotonous, relatively simple, or require continuous attention with little feedback (Gillberg & Akerstedt, 1998; Harrison & Horne, 2000; Moorcroft, 2003; Williams, Lubin, & Goodnow, 1959). Fatigue may also negatively affect other types of task performance such as those that require quick reactions, prolonged vigilance, short-term memorization (both visual and auditory), perceptual skills, or cognitive skills (Harrison & Horne, 2000; Krueger, 1989). In contrast to the job tasks mentioned above, short-term sleep debt is less likely to affect performance on tasks that are short, rule-based, or well

practiced. Minimal effects of sleepiness on tasks in which the person controls the pace, tasks that possess high intrinsic interest, or tasks that involve externally motivating rewards have also been observed (Moorcroft, 2003).

Some research concerning sleep deprivation and performance has been conducted with medical professionals. Sleep-deprived medical interns have shown hesitancy in decision-making, lack of focus when planning, lack of innovation, and impaired verbal fluency, although their ability to grasp technical information from medical journals was unaffected by sleep loss (see the review by Harrison & Horne, 2000). Similar negative effects on ability to answer medical questions and confidence level in the quality of performance were also observed among junior doctors (Lewis, Blagrove, & Ebden, 2002).

Laboratory studies have shown that sleep deprivation may result in performance decrements even when as little as two hours of sleep have been lost (Rosekind et al., 1995a; Roth, Roehrs, & Zorick, 1982). Blagrove and Akehurst (2001) found that participants deprived of sleep for 29–35 hours exhibited performance decrements on a logical reasoning task. Other findings have demonstrated that the average amount of time needed to complete perceptual and cognitive tasks increases exponentially with sleep deprivation (Babkoff et al., 1985). In addition, those tasks that initially required the most time were the most affected by the lack of sleep. Sleep deprivation for one night resulted in decreased attentiveness and subsequent hindrance of performance on both a series of inactive tasks such as monitoring warning lights and active tasks such as problem-solving (Mertens & Collins, 1986). An additional finding from this study was that simulated high altitude reduced performance when sleep deprivation was present, which has significant implications for the aviation industry.

Generally, it seems that the effects of sleep deprivation on complex physical tasks may be minimal when compared with cognitive tasks or monotonous psychomotor tasks. A sleep-deprived group and a control group performed equally well on a series of physical tasks involving muscle and anaerobic abilities (e.g., carrying sandbags), though the sleep-deprived group did experience cardiovascular deterioration over the course of the study (Rodgers et al., 1995). No differences in performance were observed between sleep-deprived participants and a control group on a complex physical task (i.e., determining a reasonable amount to lift and lifting: Legg & Haslam, 1984).

A question considered often in the early years of sleep research was the length of time needed to pass while working on a task before the effects of sleep loss were perceptible. Studies examining this question have obtained varying results, with length of time ranging from five minutes to forty minutes; however, these differences might be attributed to the type of task as well as the level of stimulation between performance sessions (Lisper & Kjellberg, 1972; Wilkinson, 1960). More recently, Gillberg and Akerstedt (1998) examined this research question in the specific context of a prolonged

vigilance task that required continuous attention. They found that performance decrements were evident after engaging in the task for only 5–10 minutes while being deprived of sleep for about 24 hours. These results suggest that the effects of a sleepless night on the performance of a monotonous task may be practically instantaneous.

Obviously the findings of the majority of these laboratory studies may not generalize easily to the common workplace, because employees are most likely not deprived of sleep frequently or for long periods of time. However, it has been shown that the tasks most affected by sleepiness are those often conducted in the workplace (e.g., cognitive, perceptual, logical) and that performance decrements occur after shorter amounts of sleep deprivation compared with the extreme lengths considered here. Furthermore, the findings of the above studies that utilize long periods of sleep deprivation may have large implications for certain organizations such as in the case of the military.

Indeed, the military itself has conducted a voluminous amount of research on the relationship between sleep and performance. Navy sailors considered to be poor-sleepers were associated with fewer promotions, lower pay grades, and higher rates of attrition when compared to good-sleepers (Johnson & Spinweber, 1982). Participants of a training course who were sleep-deprived for five days exhibited dramatic decreases in performance on both perceptual and cognitive reasoning tasks (Bugge et al., 1979). Another military sample also demonstrated performance deficits on tasks requiring vigilance and word memorization after experiencing continuous work coupled with sleep deprivation (Englund, Ryman, Naitoh, & Hodgdon, 1985). Sleep-deprived military personnel have been shown to exhibit the following behaviors: problems keeping track of critical tasks, failure to use incoming information to update maps, delay on tasks that require immediate attention, inaccuracy and misinterpretation in communication; rigidity in problem-solving, reduction in planning ahead, and exhibition of inappropriate behaviors (Harrison & Horne, 2000). Finally, performance in a military mission simulator suggested that sleep deprivation might not affect performance or visual tasks besides some minor eyestrain symptoms (e.g., eye soreness and dryness: Quant, 1992).

Absenteeism

Results from the NSF survey (2000) revealed that one out of every seven respondents indicated that they were sometimes late for work because of sleepiness. For young adults, the result was over one in every five workers. Both short- and long-term absenteeism from work has been shown to correlate with sleep disturbances (Jacquinet-Salord et al., 1993; Spector & Jex, 1998). In contrast to the above findings, absences were found to be negatively related to complaints of sleepiness on the job (Hackett & Bycio, 1996).

Safety

The next component of job performance considered is safety. Empirical evidence supports an inverse relationship between fatigue and work safety (Bourdouxhe et al., 1999). Further findings regarding the relationship between sleep and safety will be delineated in two subsections, (1) errors and (2) injuries and accidents.

(1) *Errors*. Work errors resulting from sleepiness on the job not only affect incumbents but also impact other stakeholders. It has been estimated that about 65% of human-error-caused catastrophes in the world (such as Chernobyl, Three Mile Island, and Exxon Valdez) occurred between midnight and 6 a.m., and human error is the cause of 60% to 90% of all industrial and transport accidents. In the 2000 NSF survey, approximately 20% of workers reported sometimes making mistakes at work due to sleepiness.

The total cost of medical errors has been estimated to be between $37.6 billion each year, with $17 billion of these costs associated with preventable errors. Furthermore, between 44,000 and 98,000 people in the USA die annually as a result of these errors (Kohn, Corrigan, & Donaldson, 2000). Sleepiness on the job, especially among medical residents, is thought to be a major contributing factor to the occurrence of these errors; however, medical residents are not the only group within hospitals and medical care facilities that commit errors. It has also been shown that nurses working rotating schedules report more medication errors than those that work day or evening shifts (Gold et al., 1992). As seen in the statistics cited above, these errors have major legal as well as financial implications for patient care. The impact of such errors is profound for all stakeholders, including patients, families, insurance companies, and medical service providers.

(2) *Injuries and accidents*. Two large-scale studies offer evidence about the relationship between sleep and accidents. Coren (1996) examined archive data of accidental deaths recorded by the National Center for Health Statistics and revealed that accidental deaths increased dramatically immediately following the spring shift of Daylight Savings Time. No increase was detected during the fall shift, which is consistent with the logic that the spring shift requires the loss of one hour of sleep, while the fall shift provides an extra hour of sleep. A prospective study conducted by Akerstedt, Fredlund, Gillberg, and Jansson (2002) linked phone interviews about work and health with fatal occupational accidents using the cause of death register 20 years later. The authors found that self-reported sleep problems were a predictor of accidental death at work.

The predominant amount of literature concerning accidents is in the driving context. Conservative estimates by the National Highway Traffic Safety Administration (NHTSA, 2002a) suggest that drowsy driving causes more than 100,000 crashes a year, resulting in 40,000 injuries and 1,550 deaths. More than half of US drivers reported feeling drowsy, and

20–30% reported falling asleep at the wheel. Raggatt (1991) found that number of accidents was negatively related to both sleep quantity and sleep quality.

Lyznicki, Doege, Davis, and Williams (1998) indicated that shift-workers and commercial truck-drivers were highly susceptible to having driving accidents. Long-haul truck-drivers in the US are prone to sleep deprivation because of the long hours spent driving and consequently short sleep duration (Patton, Landers, & Agarwal, 2001). Indeed, estimates suggest that truck-drivers typically average only five hours of sleep per night. Mitler, Miller, Lipsitz, Walsh, & Wylie (1997) demonstrated that over half of truck-drivers, who were videotaped and had their brain waves recorded while driving, reported feeling drowsy, and a few actually fell asleep. Fatigue may negatively impact other professional drivers besides long-haul truck-drivers (Sluiter et al., 1999). A survey of postal drivers showed that while daytime sleepiness was marginally related to driving accidents, the relationship was much stronger when only those accidents in which the driver was liable were considered (Maycock, 1997).

Medical residents, one profession characterized by erratic schedules involving many shifts, frequently fall asleep when driving after work (Patton et al., 2001), and are nearly seven times more likely to have an accident than before they started their residencies. Nurses working nights or rotating shifts were more likely to report nodding off during driving to or from work and experiencing 'near-miss' automobile accidents than those working day or evening shifts (Gold et al., 1992).

Although lack of sufficient sleep plays a major role in sleep-related vehicular accidents, the time of day is also important. The peak occurrences of accidents resulting from sleepiness are around 2 a.m., 6 a.m., and 4 p.m. (Horne & Reyner, 1995). These are the times of peak circadian sleepiness shown in Figure 3.1. When the number of people driving at these different times of the day are taken into account, the risk of a sleep-related accident at 6 a.m. is 20 times greater and at 4 p.m. three times greater than at 10 a.m. (Moorcroft, 2003).

Empirical evidence supporting the relationship between sleep and occupational accidents besides those involving transportation is minimal, even though the connection may be obvious from a logic perspective. The data used for Parkes' (1999) study showed no relationship between sleep problems and work-related injuries (K. R. Parkes, personal communication, June 17, 2002). In contrast, in a prospective study, men who reported both excessive daytime sleepiness and snoring at baseline were at an increased risk of occupational accidents during the following 10 years, with an odds ratio of 2.2, while controlling for factors such as age, body mass index, smoking, alcohol dependence, work tenure, blue-collar job, shift work, and exposure to noise, organic solvents, exhaust fumes, and whole-body vibrations (Lindberg, Carter, Gislason, & Janson, 2001). The authors also reported a 95% con-

fidence interval of 1.3–3.8 associated with the odds ratio, suggesting that a significant relationship did indeed exist between sleep symptoms and work-related accidents.

Creativity

Horne (1988) examined the differences in 'divergent' thinking ability or creativity between subjects deprived of sleep for over 24 hours and control subjects who were not sleep-deprived. The results showed that multiple dimensions of creativity including flexibility and originality were significantly impaired in participants experiencing sleep deprivation compared with those who were not. No effects of sleep debt were detected on 'convergent' thinking tasks or those not requiring creativity. Another study did not find a significant relationship between sleep and creativity, though sleep was not manipulated and a different measure of creativity was used (Narayanan, Vijayakumar, & Govindarasu, 1992). Lewin and Glaubman (1975) also demonstrated that subjects showed decreased levels of creativity on some tasks when their REM sleep was deprived. An explanatory factor for decreased creative performance may be that people who suffer from sleep deprivation tend to be more susceptible to argument and suggestion and less capable of anticipating ranges of possible consequences (Harrison & Horne, 2000).

Although there are minimal empirical studies that directly examine the antecedents and consequences of sleepiness on the job, indirect evidence from the above review offers some insight into the development of preventive as well as maintenance strategies to cope with workplace sleepiness.

COUNTERMEASURES OF SLEEPINESS IN THE WORKPLACE

Given the complexity of sleep, the plausible impacts of task characteristics and job contexts, and differences among individuals, it is unrealistic and impractical to develop a one-size-fits-all solution to eliminate sleepiness in the workplace. In this section, we consider two major types of countermeasure specifically focused at the individual and organizational levels. While the former emphasizes approaches that employees can utilize to counter their sleepiness at work, the latter focuses on strategies organizations can implement to improve quality of work.

Individual Countermeasures

Countermeasures initiated by employees exist in two forms: (1) things that can be done before work and during rest periods and (2) things that can be done while at work (Rosekind et al., 1996a, b). For most people, sleepiness that is not associated with a sleep disorder can be alleviated by these countermeasures.

Minimizing sleep debt

Because the negative effects of sleep loss increase exponentially, employees who experience sleep debt due to work demands (e.g., intense and frequent sales trips, urgency of debugging a computer program) should catch up on lost sleep as soon as possible. Ideally, those sleep-deprived should have two nights of unrestricted sleep, which requires latitude to sleep when drowsiness is experienced and rise when natural awakening occurs. Unfortunately, it seems that people are not able to sleep prophylactically in anticipation of future sleep debt resulting from heightened job demands (e.g., a pressing deadline).

Good sleep hygiene

To make sleep onset come quickly as well as easily, employees should maintain good sleep hygiene as described below (Moorcroft, 2003):

1 Go to bed at the same time and wake up at the same time every day, although rising at the same time is the more imperative of the two. For instance, try to get up at your normal time even after working long hours the day before.
2 Choose the times to go to bed and wake up so that you get approximately eight hours of sleep per night.
3 Arrange the bedroom to make sleeping easier. Specifically, a comfortable bed, a dark and quiet bedroom at a comfortable temperature (cooler is better than warmer), and some humidity can facilitate sleep. Refrain from engaging in other activities (e.g., watching television, reading) besides sleep while in bed.
4 Have a pre-sleep routine that is calming and provides a separation between the sleeping and waking parts of your day. This routine might include bathing, teeth cleaning, reading, or meditating.
5 Avoid rigorous activities and aerobic exercise in the hours directly prior to bedtime. Regular exercise several hours before bedtime may actually increase sleep quality.
6 Refrain from drinking alcohol, eating or drinking caffeine, and smoking cigarettes several hours before bedtime. Although alcohol is a depressant and may cause sleepiness, it often results in fitful sleep throughout the night.

Diet

Empirical research has substantiated the beneficial effects of certain diets. For instance, studies conducted in an interactive driving simulator showed that a glucose-based 'energy' drink significantly improved sleepy drivers' lane drifting and reaction time (Horne & Reyner, 2001), as well as reduced

sleep-related driving incidents and subjective and objective (EEG readings) sleepiness (Reyner & Horne, 2002). Beyond these positive findings associated with the energy drink, evidence of the positive effects of specific types of food on alertness and performance is not conclusive. For instance, foods rich in carbohydrates may induce sleep after a transient alertness. In contrast, foods high in protein are proposed to promote wakefulness (Rosekind et al., 1995a).

Working the biological clock

People whose jobs require them to work at times other than during the day or to travel rapidly across several time zones have sleep problems because of disruptions to their biological clocks. Unfortunately, no especially effective remedies are available to counteract these disruptions, but some measures can be taken in order to keep the harmful effects at a minimum. An individual's biological clock can be reset by controlled exposure to sunlight and engagement in a social routine at specific times for several days but this is often not feasible. Part of the problem is that there is no simple, accurate way to read one's biological clock and thus determine when the sunlight exposure and social routine should occur. If these activities are completed at the wrong times, the result can be no effect or even counterproductive. Current research is investigating how to use light, activity, and drugs (e.g., melatonin) among other things to help individuals practically and effectively reset their biological clocks as required by their job.

Drugs

A beneficial effect on alertness is observed about 30 minutes after 150 mg of caffeine is consumed. Recent studies are proving the positive effects of slowly released caffeine (SRC) on performance and alertness during 9–13-hr periods with no major side-effects. A single daily dose of 600 mg or two 300-mg doses of SRC are shown to have better effects than repeated doses of caffeine that range from moderate to high potency on performance and alertness for periods of wakefulness consisting of 24 hours or longer (Beaumont et al., 2001). Similar to caffeine, modafinil also maintains alertness levels during the morning hours after sleep deprivation. Studies show that modafinil does not appear to offer any certain advantages over the effects of caffeine for improving alertness and performance levels by healthy adults after significant sleep debt (Wesensten et al., 2001).

Naps

As shown in Figure 3.1, alertness level tends to decrease in the mid-afternoon. Hence, a short afternoon nap can be highly effective in combating sleepiness (Seo et al., 2000); however, longer naps may cause sleep inertia

(Moorcroft, 2003). Empirical research has shown that a 20-minute nap in the mid-afternoon improved subjective sleepiness, performance levels, and confidence in task performance (Hayashi, Watanabe, & Hori, 1999). The positive effects associated with an afternoon nap have also been observed for people experiencing sleep debt (Takahashi & Arito, 2000). Compared with an afternoon nap, a noontime nap has only partial positive effects (Hayashi, Ito, & Hori, 1999). Note that the combination of drinking coffee followed immediately by a 20-minute nap has been shown to provide positive effects that endure for several hours, which significantly reduces the likelihood of having a sleep-related accident (Moorcroft, 2003).

In addition to the afternoon nap, other naptimes available for employees have been shown to have positive effects. For instance, Rosekind et al. (1995a, b) documented that long-haul pilots who were allowed a 40-minute nap performed 34% better and were twice as alert than cohorts who did not nap. After a yearlong monitoring of shift-workers, Bonnefond et al. (2001) substantiated the positive effects of a short nap during the night shift. Specifically, night-workers who engaged in short naps had greater self-reported job satisfaction, vigilance, and quality of life compared with non-napping workers.

Organizational Countermeasures

Countermeasures implemented by organizations include stress management, fatigue management, education, work schedule design, workplace design, and personnel selection and placement.

Stress management

As reviewed earlier, relationships between job stressors and psychosomatic symptoms (e.g., Spector, Chen, & O'Connell, 2000) along with sleep quality (e.g., Farnill & Robertson, 1990) have been consistently replicated. Individual-oriented stress management interventions can be easily offered to employees with minimal disruption of work routines (Murphy, 1988). According to Bellarosa and Chen (1997), relaxation is the most common intervention and is viewed as the most practical compared with five other types of intervention (e.g., meditation). However, the authors note that stress management experts considered physical exercises as the most effective intervention among them all.

Fatigue management

As reviewed earlier, a short nap has been associated with positive effects (Bonnefond et al., 2001; Rosekind et al., 1995a, b). Organizations can provide both space and opportunity for planned naps especially for key

personnel whose jobs are safety-sensitive. An exemplary circumstance is that of emergency medicine doctors who are on-call for extended hours. Since these doctors are often required to remain at the hospital while on-call, they are encouraged to nap in between call-ins. In reality, the availability of naps has been implemented by only a handful of businesses thus far.

Education

Both incumbents and organizations can benefit from education about the interrelationships of changes in the body's circadian rhythms, sleep problems, health symptoms, and strains that result from shift work (Smith et al., 1999). Teaching shift-workers about good sleep hygiene can also be beneficial. Many educational materials are freely disseminated by federal (e.g., National Highway Traffic Safety Administration or NHTSA) and private agencies (e.g., National Sleep Foundation). These materials describe the basics of sleep including circadian factors, the biological basis of sleepiness, misconceptions about sleep and sleepiness, and countermeasures relevant to jobs with high propensity of sleepiness. With the aid of federal agencies, organizations can deliver training programs about dealing with drowsiness while commuting or about increasing shift work tolerance. For example, the NHTSA provides an *Employer Administrator's Guide* that includes ways to prevent drowsy driving by shift-workers, which is available at http://www.nhtsa.dot.gov/people/perform/human/drows_driving/resource/resource.html.

Work schedule design

Organizations can also establish policies concerning work hours that are consistent with the current knowledge about sleep and fatigue. Since it is known when in the 24-hr day errors and accidents due to sleepiness are more likely to occur, implementing policies and procedures to lessen the extent that the mistake-prone job tasks occur during these times would be cost-effective (Mitler & Miller, 1996). Also, rotating shift-workers clockwise has been shown to be far better than counterclockwise (Akerstedt, 1995; Maas et al., 1999). Mass et al. reported that an oil refinery saved almost \$2$\frac{1}{2}$ million by implementing a better shift schedule because of less overtime, less idle time, reduced absences, and better worker health and safety. In general, a shift work schedule should not require more than five consecutive nights, more than four consecutive 12-hr shifts, or a day shift start time earlier than 7 a.m. Complicated schedules should also be avoided. For more information about the ideal characteristics of a shift work schedule (e.g., shift length, shift start and end time, shift extension or doubling, opportunity to rest prior to work, opportunity to recover, number of consecutive shifts), refer to Smith, Folkard, and Fuller (2002).

Workplace design

Organizations can design the workplace and adopt technological innovations to minimize sleepiness, especially for jobs that involve high levels of accident risk and are sensitive to decreased attention due to sleepiness (Mitler & Miller, 1996). For instance, the workplace should have bright lighting (more than 7,000 lux) and cool but comfortable temperatures with plenty of air changes per hour. In addition, employees should have easy access to healthy food, which can counter sleepiness.

Personnel selection or placement

Traditional personnel selection or placement decisions are made based upon the prediction that an applicant will be more satisfactory than other applicants or an employee will be more satisfactory in one position than another position, respectively. If certain personal characteristics are deemed job-related and substantial validity evidence exists pertaining to the relationships between these personal characteristics and shift work performance (e.g., adjustment, task performance), these personal characteristics may be considered during selection and placement decisions.

However, as demonstrated in the review of individual antecedents, conclusions about these relationships are often tenuous at best. For instance, empirical evidence of the relationship between evening type (i.e., phase of the circadian rhythm) and shift work tolerance tends to be weak or moderate (Bohle & Tilley, 1989; Steele, Ma, Watson, & Thomas, 2000), with some exceptions (Costa, Lievore, Casaletti, Gaffuri, & Folkard, 1989; Gander, Nguyen, Rosekind, & Connell, 1993; Kaliterna, Vidacek, Prizmic, & Radosevic-Vidacek, 1995). On the other hand, the stability and the amplitude of the circadian rhythm tend to predict shift work tolerance in a more consistent fashion (Costa et al., 1989; Vidacek, Kaliterna, & Radosevic-Vidacek, 1987; Steele et al., 2000). Vidacek, Radosevic-Vidacek, Kaliterna, & Prizmic (1993) also revealed that shift-workers who had higher positive moods, lower negative moods, and lower fatigue prior to their work tended to show shift work tolerance. Until more systematic research is conducted regarding the potential predictability of these individual antecedents for performance affected by sleepiness, consideration of these individual characteristics during selection and placement decisions will be limited.

CONCLUSION

To most organizations, workplace sleepiness is most often considered a problem needed to be dealt with by the individual employee. However, according to the current review, we argue that sleepiness on the job is an

epidemic, occupational health problem that requires attention from various stakeholders including practitioners, policy-makers, as well as researchers. The costs associated with consequences of sleepiness on the job can be astronomical (litigation, accidents, productivity, health care, etc.), and the impacts can be pervasive across families, communities, organizations, and societies as a whole (Mitler, Dement, & Dinges, 2000; Sparks et al., 1997).

Because only limited research pertaining to sleepiness in the workplace has been conducted in I/O psychology, we applied the facet analysis approach to generate plausible antecedents and consequences facets, which can guide the delineation of plausible causal relationships among the facets. In general, the literature suggests sleepiness in the workplace is prevalent. Although there are some inconclusive results about the effects of workplace sleepiness, its relationships with psychological as well as physical well-being (Barton et al., 1995; Pilcher & Ott, 1998), performance (Harrison & Horne, 2000), and safety (Lindberg et al., 2001) were consistently substantiated.

Our review further suggested that the level of workplace sleepiness might vary contingent upon demographics (e.g., age, Parkes, 1994), personality (e.g., neuroticism, Blagrove & Akehurst, 2001), circadian rhythm (Khaleque, 1999), type of task (Gander et al., 1998), work schedule (Akerstedt, 1990), work environment (Parkes, 2002), or type of job stressor (Spector & Jex, 1998). Only with systematic research in the future will these scientific inquiries be confirmed.

REFERENCES

Akerstedt, T. (1990). Psychological and psychophysiological effects of shift work. *Scandinavian Journal of Work, Environment & Health*, **16**, 67–73.

Akerstedt, T. (1995). Work hours, sleepiness and accidents: Introduction and summary. *Journal of Sleep Research*, **4**, 1–3.

Akerstedt, T., Fredlund, P., Gillberg, M., & Jansson, B. (2002). A prospective study of fatal occupational accidents: Relationship to sleeping difficulties and occupational factors. *Journal of Sleep Research*, **11**, 69–71.

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, **52**, 29–37.

Akerstedt, T., & Torsvall, L. (1981). Shift work: Shift-dependent well-being and individual differences. *Ergonomics*, **24**, 265–273.

Alapin, I., Fichten, C. S., Libman, E., Creti, L., Bailes, S., & Wright, J. (2000). How is good and poor sleep in older adults and college students related to daytime sleepiness, fatigue, and ability to concentrate? *Journal of Psychosomatic Research*, **49**, 381–390.

Appels, A., & Schouten, E. (1991). Waking up exhausted as a risk indicator of myocardial exhaustion. *American Journal of Cardiology*, **68**, 395–398.

Babkoff, H., Thorne, D. R., Sing, H. C., Genser, S. G., et al. (1985). Dynamic changes in work/rest duty cycles in a study of sleep deprivation. *Behavior Research Methods, Instruments, & Computers*, **17**, 604–613.

Bacharach, S. B., Bamberger, P., & Conley, S. (1991). Workhome conflict among nurses and engineers: Mediating the impact of role stress on burnout and satisfaction at work. *Journal of Organizational Behavior*, **12**, 39–53.

Barton, J., Spelten, E., Totterdell, P., Smith, L., Folkard, S., & Costa, G. (1995). The Standard Shiftwork Index: A battery of questionnaires for assessing shiftwork-related problems. *Work and Stress*, **9**, 4–30.

Beaumont, M., Batejat, D., Pierard, C., Coste, O., Doireau, P., Van Beers, P., et al. (2001). Slow release caffeine and prolonged (64-h) continuous wakefulness: Effects on vigilance and cognitive performance. *Journal of Sleep Research*, **10**, 265–276.

Beehr, T. A., & Newman, J. E. (1978). Job stress, employee health, and organizational effectiveness: A facet analysis, model, and literature review. *Personnel Psychology*, **31**, 665–699.

Beermann, B., & Nachreiner, F. (1995). Working shifts—different effects for women and men? *Work & Stress*, **9**, 289–297.

Bellarosa, C., & Chen, P. Y. (1997). The effectiveness and practicality of occupational stress management interventions: A survey of subject matter expert opinions. *Journal of Occupational Health Psychology*, **2**, 247–262.

Benca, R., & Kwapil, T. (2000). Sleep disorders measures. In A. J. Rush and others (eds), *Handbook of Psychiatric Measures* (pp. 673–685). Washington, DC: American Psychiatric Association.

Bergmann, T. J., & Volkema, R. J. (1994). Issues, behavioral responses and consequences in interpersonal conflicts. *Journal of Organizational Behavior*, **15**, 467–471.

Berry, D. T. R., & Webb, W. B. (1985). Mood and sleep in aging women. *Journal of Personality and Social Psychology*, **49**, 1724–1727.

Blagrove, M., & Akehurst, L. (2001). Personality and the modulation of effects of sleep loss on mood and cognition. *Personality & Individual Differences*, **30**, 819–828.

Bohle, P., & Tilley, A. J. (1989). The impact of night work on psychological well-being. *Ergonomics*, **32**, 1089–1099.

Bohle, P., & Tilley, A. J. (1993). Predicting mood change in night shift. *Ergonomics*, **36**, 125–133.

Boles, J. S., Johnston, M. W., & Hair, J. F., Jr (1997). Role stress, work–family conflict and emotional exhaustion: Inter-relationships and effects on some work-related consequences. *Journal of Personal Selling & Sales Management*, **17**, 17–28.

Bonnefond, A., Muzet, A., Winter-Dill, A., Bailloeuil, C., Bitouze, F., & Bonneau, A. (2001). Innovative working schedule: Introducing one short nap during the night shift. *Ergonomics*, **44**, 937–945.

Bourdouxhe, M. A., Queinnec, Y., Granger, D., Baril, R. H., Guertin, S. C., Massicotte, P. R., et al. (1999). Aging and shiftwork: The effects of 20 years of rotating 12-hour shifts among petroleum refinery operators. *Experimental Aging Research*, **25**, 323–329.

Breaugh, J. A. (1983). The 12-hour workday: Differing employee reactions. *Personnel Psychology*, **36**, 277–288.

Breslau, N., Roth, T., Rosenthal L., & Andreski, P. (1997). Daytime sleepiness: An epidemiological study of young adults. *American Journal of Public Health*, **87**, 1649–1653.

Bugge, J. F., Opstad, P. K., & Magnus, P. M. (1979). Changes in the circadian rhythm of performance and mood in healthy young men exposed to prolonged, heavy physical work, sleep deprivation, and caloric deficit. *Aviation Space & Environmental Medicine*, **50**, 663–668.

Burke, R. J. (1994). Stressful events, work–family conflict, coping, psychological burnout, and well-being among police officers. *Psychological Reports*, **75**, 787–800.

Buysse, D. J., Reynolds, C. F., III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, **28**, 193–213.

Caldwell, J. A., Jr, & LeDuc, P. A. (1998). Gender influences on performance, mood and recovery sleep in fatigued aviators. *Ergonomics*, **41**, 1757–1770.

Carskadon, M. A., & Dement, W. C. (2000). Normal human sleep: An overview. In M. H. Kryger, T. Roth, & W. C. Dement (eds), *Principles and Practice of Sleep Medicine* (3rd edn, pp. 15–25). Philadelphia: Saunders.

CBS (2002, May 30). Towboat captain was short on sleep. CBSNEWS.com. Retrieved May 31, 2002, from http://www.cbsnews.com/stories/2002/05/26/national/main510195.shtml. Columbia Broadcasting System.

Chervin, R. D., Aldrich, M. S., Pickett, R., & Guilleminault, C. (1997). Comparison of the results of the Epworth Sleepiness Scale and the multiple sleep latency test. *Journal of Psychosomatic Research*, **42**, 145–155.

Colquhoun, W. P. (1971). *Biological Rhythms and Human Performance*. London: Academic Press.

Coren, S. (1996). *Sleep Thieves*. New York: Free Press.

Costa, G., Lievore, F., Casaletti, G., Gaffuri, E., & Folkard, S. (1989). Circadian characteristics influencing interindividual differences in tolerance and adjustment to shiftwork. *Ergonomics*, **3**, 373–385.

Culebras, A. (2002). Normal sleep. In T. L. Lee-Chiong, Jr, M. J. Sateia, & M. A. Carskadon (eds), *Sleep Medicine* (pp. 1–6). Philadelphia: Hanley & Belfus.

Czeisler, C. A., Moore-Ede, M. C., & Coleman, R. M. (1982). Rotating shift work schedules that disrupt sleep are improved by applying circadian principles. *Science*, **217**, 460–463.

Dalbokova, D., Tzenova, B., & Ognjanova, V. (1995). Stress states in nuclear operators under conditions of shiftwork. *Work & Stress*, **9**, 305–313.

Danker-Hopfe, H., Kraemer, S., Dorn, H., Schmidt, A., Ehlert, I., & Herrmann, W. M. (2001). Time-of-day variations in different measures of sleepiness (MSLT, pupillography, and SSS) and their interrelations. *Psychophysiology*, **38**, 828–835.

Dawson, D., & Fletcher, A. (2001). A quantitative model of work-related fatigue: Background and definition. *Ergonomics*, **44**, 144–163.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., et al. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep*, **20**, 267–277.

Donaldson, S. I., Sussman, S., Dent, C. W., Severson, H. H., & Stoddard, J. L. (1999). Health behavior, quality of work life, and organizational effectiveness in the lumber industry. *Health Education & Behavior*, **26**, 579–591.

Drummond, S. P., & Brown, G. G. (2001). The effects of total sleep deprivation on cerebral responses to cognitive performance. *Neuropsychopharmacology*, **25**, 568–573.

Ellis, B. W., Johns, M. W., Lancaster, R., Raptopoulos, P., Angelopoulos, N., & Priest, R. G. (1981). The St. Mary's Hospital Sleep Questionnaire: A study of reliability. *Sleep*, **4**, 93–97.

Englund, C. E., Ryman, D. H., Naitoh, P., & Hodgdon, J. A. (1985). Cognitive performance during successive sustained physical work episodes. *Behavior Research Methods, Instruments, & Computers*, **17**, 75–85.

Farnill, D., & Robertson, M. F. (1990). Sleep disturbance, tertiary-transition stress, and psychological symptoms among young first-year Australian college students. *Australian Psychologist*, **25**, 178–188.

Folkard, S., & Monk, T. H. (1979). Shiftwork and performance. *Human Factors*, **21**, 483–492.

Folkard, S., Monk, T. H., & Lobban, M. (1979). Towards a predictive test of adjustment to shiftwork. *Ergonomics*, **22**, 79–91.

Folstein, M. F., & Luria, R. (1973). Reliability, validity, and clinical application of the Visual Analogue Mood Scale. *Psychological Medicine*, **3**, 479–486.

Freyd, M. (1923). The graphic rating scale. *Journal of Educational Psychology*, **14**, 83–102.

Frost, P. J., & Jamal, M. (1979). Shift work, attitudes, and reported behavior: Some associations between individual characteristics and hours of work and leisure. *Journal of Applied Psychology*, **64**, 77–81.

Gander, P. H., Nguyen, D, Rosekind, M. R., & Connell, L. J. (1993). Age, circadian-rhythms, and sleep loss in flight crews. *Aviation Space and Environmental Medicine*, **64**, 189–195.

Gander, P. H., Rosekind, M. R., & Gregory, K. B. (1998). Flight crew fatigue VI: A synthesis. *Aviation Space & Environmental Medicine*, **69**, B49–B60.

Garbarino, S., Beelke, M., Costa, G., Violani, C., Lucidi, F., Ferrillo, F., et al. (2002). Brain function and effects of shift work: Implications for clinical neuropharmacology. *Neuropsychobiology*, **45**, 50–56.

Gaudreau, H., Morettini, J., Lavoie, H. B., & Carrier, J. (2001). Effects of a 25-h sleep deprivation on daytime sleep in the middle-aged. *Neurobiology of Aging*, **22**, 461–468.

Gillberg, M., & Akerstedt, T. (1998). Sleep loss performance: No 'safe' duration of a monotonous task. *Physiology & Behavior*, **64**, 599–604.

Gillberg, M., Kecklund, G., & Akerstedt, T. (1994). Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep*, **17**, 236–241.

Gold, D. R., Rogacz, S., Bock, N., Tosteson, T. D, Baum, T. M., Speizer, F. E., et al. (1992). Rotating shift work, sleep, and accidents related to sleepiness in hospital nurses. *American Journal of Public Health*, **82**, 1011–1014.

Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, **70**, 177–206.

Guilleminault, C., & Poyares, D. (2001). Violence during sleep. Retrieved May 31, 2002, from http://www.aasmnet.org/MEDSleep/violenceslides.htm.

Hackett, R. D., & Bycio, P. (1996). An evaluation of employee absenteeism as a coping mechanism among hospital nurses. *Journal of Occupational and Organizational Psychology*, **69**, 327–338.

Haermae, M. (1993). Individual differences in tolerance to shiftwork: A review. *Ergonomics*, **13**, 101–109.

Haermae, M., Tenkanen, L., Sjoeblom, T., Alikoski, T., & Heinsalmi, P. (1998). Combined effects of shift work and life-style on the prevalence of insomnia, sleep deprivation and daytime sleepiness. *Scandinavian Journal of Work, Environment & Health*, **24**, 300–307.

Harrington, J. M. (1994). Shift work and health—a critical review of the literature on working hours. *Annals Academy Medicine*, **23**, 699–705.

Harrison, Y., & Horne, J. A. (1997). Sleep deprivation affects speech. *Sleep*, **20**, 871–877.

Harrison, Y., & Horne, J. A. (1999). One night of sleep loss impairs innovative thinking and flexible decision making. *Organizational Behavior and Human Decision Processes*, **78**, 128–145.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology-Applied*, **6**, 236–249.

Hayashi, M., Ito, S., & Hori, T. (1999). The effects of a 20-min nap at noon on sleepiness, performance and EEG activity. *International Journal of Psychophysiology*, **32**, 173–180.

Hayashi, M., Watanabe, M., & Hori, T. (1999). The effects of a 20 min nap in the mid-afternoon on mood, performance and EEG activity. *Clinical Neurophysiology*, **110**, 272–279.

Hicks, R. A., & Pellegrini, R.J. (1978). Locus of control in short and long sleepers. *Perceptual & Motor Skills*, **47**, 1337–1338.

Hildebrandt, G., & Stratmann, I. (1979). Circadian system response to night work in relation to the individual circadian phase position. *International Archives of Occupational and Environmental Health*, **43**, 73–83.

Hillhouse, J. J., Adler, C. M., & Walters, D. N. (2000). A simple model of stress, burnout, and symptomatology in medical residents: A longitudinal study. *Psychology Health & Medicine*, **5**, 63–73.

Hoddes, E., Dement, W. C., & Zarcone, V. (1972). The history and use of the Stanford Sleepiness Scale. *Psychophysiology*, **9**, 150.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973) Quantification of sleepiness: A new approach. *Psychophysiology*, **10**, 431–436.

Hoogendoorn, W. E., Bongers, P. M., de Vet, H. C. W., Houtman, I. L. D., Ariens, G. A. M., van Mechelen, W., & Bouter, L. M. (2001). Psychosocial work characteristics and psychological strain in relation to low-back pain. *Scandinavian Journal of Work, Environment & Health*, **28**, 258–267.

Horne, J. A. (1988). Sleep loss and 'divergent' thinking ability. *Sleep*, **11**, 528–536.

Horne, J. A., & Reyner, L. A. (1995). Sleep-related vehicle accidents. *British Medical Journal*, **310**, 565–567.

Horne, J. A., & Reyner, L. A. (2001). Beneficial effects of an 'energy drink' given to sleepy drivers. *Amino Acids*, **20**, 83–89.

Hurrell, J. J., & Colligan, M. J. (1986). Machine pacing and shiftwork: Evidence for job stress. *Journal of Organizational Behavior Management*, **8**, 159–175.

Jacquinet-Salord, M. C., Lang, T., Fouriaud, C., Nicoulet, I., & Bingham, A. (1993). Sleeping tablet consumption, self-reported quality of sleep, and working conditions. *Journal of Epidemiology and Community Health*, **47**, 64–68.

Jean-Louis, G., Kripke, D. F., & Ancoli-Israel, S. (2000). Sleep and quality of well-being. *Sleep*, **23**, 1115–1119.

Jean-Louis, G., Kripke, D. F., Assmus, J. D., & Langer, R. D. (2000). Sleep duration, illumination, and activity patterns in a population sample: Effects of gender and ethnicity. *Biological Psychiatry*, **47**, M120–M123.

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. *Sleep*, **14**, 540–545.

Johns, M. W. (1992). Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, **15**, 376–381.

Johns, M. W. (1994). Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep*, **17**, 703–710.

Johnson, L. C., Freeman, C. R., Spinweber, C. L., & Gomez, S. A. (1991). Subjective and objective measures of sleepiness: Effect of benzodiazepine and caffeine on their relationship. *Psychophysiology*, **28**, 65–71.

Johnson, L. C., & Spinweber, C. L. (1982). Quality of sleep and performance in the Navy: A longitudinal study of good and poor sleepers. *US Naval Health Research Center Report*, **82**(8), 1–14.
Kalimo, R., Tenkanen, L., Haermae, M., Poppius, E., & Heinsalmi, P. (2000). Job stress and sleep disorders: Findings from the Helsinki Heart Study. *Stress Medicine*, **16**, 65–75.
Kaliterna, L., Vidacek, S., Prizmic, Z., & Radosevic-Vidacek, B. (1995). Is tolerance to shiftwork predictable from individual difference measures? *Work & Stress*, **9**, 140–147.
Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, **24**, 285–311.
Khaleque, A. (1999). Sleep deficiency and quality of life of shift workers. *Social Indicators Research*, **46**, 181–189.
Kogi, K. (1991). Job content and working time: The scope for joint change. *Ergonomics*, **34**, 757–773.
Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press.
Koulack, D., & Nesca, M. (1992). Sleep parameters of Type A and B scoring college students. *Perceptual & Motor Skills*, **74**, 723–726.
Kramer, M., Roehrs, T., & Roth, T. (1976). Mood change and the physiology of sleep. *Comprehensive Psychiatry*, **17**, 161–165.
Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work & Stress*, **3**, 129–141.
Kumar, A., & Vaidya, A. K. (1982). Neuroticism in short and long sleepers. *Perceptual & Motor Skills*, **54**, 962.
Kumar, A., & Vaidya, A. K. (1986). Locus of control in short and long sleepers. *British Journal of Psychiatry*, **148**, 739–740.
Landis, C. A., Bergmann, B. M., Ismail, M. M., & Rechtschaffen, A. (1992). Sleep deprivation in the rat: XV. Ambient temperature choice in paradoxical sleep-deprived rats. *Sleep*, **15**, 13–20.
Lee, K. A. (1992). Self-reported sleep disturbances in employed women. *Sleep*, **15**, 493–498.
Lee, K. A., Hicks, G., & Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Research*, **36**, 291–298.
Legg, S. J., & Haslam, D. R. (1984). Effect of sleep deprivation on self-selected workload. *Ergonomics*, **27**, 389–396.
Lewin, I., & Glaubman, H. (1975). The effect of REM deprivation: Is it detrimental, beneficial, or neutral? *Psychophysiology*, **12**, 349–353.
Lewis, K. E., Blagrove, M., & Ebden, P. (2002). Sleep deprivation and junior doctors' performance and confidence. *Postgraduate Medical Journal*, **78**, 85–87.
Lilley, R., Feyer, A-M., Kirk, P., & Gander, P. (2002). A survey of forest workers in New Zealand: Do hours of work, rest, and recovery play a role in accidents and injury? *Journal of Safety Research*, **33**, 53–71.
Lindberg, E., Carter, N., Gislason, T., & Janson, C. (2001). Role of snoring and daytime sleepiness in occupational accidents. *American Journal of Respiratory & Critical Care Medicine*, **164**, 2031–2035.
Lindstroem, K., & Hurrell, J. J. (1992). Coping with job stress by managers at different career stages in Finland and the United States. *Scandinavian Journal of Work, Environment & Health*, **18**, 14–17.
Lingenfelser, T., Kaschel, R., Weber, A., Zaiser-Kaschel, H., Jakober, B., & Kuper, J. (1994). Young hospital doctors after night duty: Their task-specific cognitive status and emotional condition. *Medical Education*, **28**, 566–572.

Lisper, H. O., & Kjellberg, A. (1972). Effects of 24-hour sleep deprivations on rate of decrement in a 10-minute auditory reaction time task. *Journal of Experimental Psychology*, **96**, 287–290.

Loewenthal, K. M., Eysenck, M., Harris, D., Lubitsh, G., Gorton, T., & Bicknell, H. (2000). Stress, distress, and air traffic incidents: Job dysfunction and distress in airline pilots in relation to contextually-assessed stress. *Stress Medicine*, **16**, 179-183.

Lowden, A., Kecklund, G., Axelsson, J., & Akerstedt, T. (1998). Change from an 8-hour shift to a 12-hour shift, attitudes, sleep, sleepiness and performance. *Scandinavian Journal of Work, Environment & Health*, **24**, 69–75.

Luna, T. D., French, J., & Mitcha, J. L. (1997). A study of USAF air traffic controller shiftwork: Sleep, fatigue, activity, and mood analyses. *Aviation Space & Environmental Medicine*, **68**, 18–23.

Lyznicki, J. M., Doege, T. C., Davis, R. M., & Williams, M. A. (1998). Sleepiness, driving, and motor vehicle crashes. *JAMA: Journal of the American Medical Association*, **279**, 1908–1913.

Maas, J. B., Axelrod, D. J., & Hogan, B. R. (1999). *Power Sleep: The Revolutionary Program that Prepares Your Mind for Peak Performance*. New York: HarperTrade.

Marquie, J. C., Foret, J., & Queinnec, Y. (1999). Effects of age, working hours, and job content on sleep: A pilot study. *Experimental Aging Research*, **25**, 421–427.

Martens, M. F. J., Nijhuis, F. J. N., Van Boxtel, M. P. J., & Knottnerus, J. A. (1999). Flexible work schedules and mental and physical health. A study of a working population with non-traditional working hours. *Journal of Organizational Behavior*, **20**, 35–46.

Maycock, G. (1997). Sleepiness and driving: The experience of U.K. car drivers. *Accident Analysis & Prevention*, **29**, 453–462.

Mecacci, L., & Zani, A. (1983). Morningness–eveningness preferences and sleep–waking diary data of morning and evening types in student and worker samples. *Ergonomics*, **26**, 1147–1153.

Mertens, H. W., & Collins, W. E. (1986). The effects of age, sleep deprivation, and altitude on complex performance. *Human Factors*, **28**, 541–551.

Mitchell, R. J., & Williamson, A. M. (2000). Evaluation of an 8 hour versus a 12 hour shift roster on employees at a power station. *Applied Ergonomics*, **31**, 83–93.

Mitler, M. M. (2002). Forensic sleep medicine. In T. L. Lee-Chiong, Jr, M. J. Sateia, and M. A. Carskadon (eds), *Sleep Medicine* (pp. 143–150). Philadelphia: Hanley & Belfus.

Mitler, M. M., Carskadon, M. A., & Hirshkowitz, M. (2000). Evaluating sleepiness. In M. H. Kryger, T. Roth, & W. C. Dement (eds), *Principles and Practice of Sleep Medicine* (3rd edn, pp. 1251–1258). Philadelphia: Saunders.

Mitler, M. M., Dement, W. C., & Dinges, D. F. (2000). Sleep medicine, public policy, and public health. In M. H. Kryger, T. Roth, & W. C. Dement (eds), *Principles and Practice of Sleep Medicine* (3rd edn, pp. 580–588). Philadelphia: Saunders.

Mitler, M. M., Gujavarty, K. S., & Browman, C. P. (1982). Maintenance of wakefulness test: A polysomnographic technique for evaluation of treatment efficacy in patients with excessive somnolence. *Electroencephalography and Clinical Neurophysiology*, **53**, 658–661.

Mitler, M. M., & Miller, J. C. (1996). Some practical considerations and policy implications of studies of sleep patterns. *Behavioral Medicine*, **21**, 184–185.

Mitler, M. M., Miller, J. C., Lipsitz, J. J., Walsh, J. K., & Wylie, C. D. (1997). The sleep of long-haul truck drivers. *New England Journal of Medicine*, **337**, 755–761.

Monk, T. H., & Aplin, L. C. (1980). Spring and autumn daylight saving time changes: Studies of adjustment in sleep timings, mood, and efficiency. *Ergonomics*, **23**, 167–178.

Moorcroft, W. M. (1993). *Sleep, Dreaming, and Sleep Disorders: An Introduction* (2nd edn). Lanham, MD: University Press of America.

Moorcroft, W. M. (2003). *Understanding Sleep and Dreaming*. New York: Kluwer Academic Press.

Murphy, L. R. (1988). Workplace interventions for stress reduction and prevention. In C. L., Cooper & R. Payne (eds), *Causes, Coping and Consequences of Stress at Work* (pp. 301–339). Chichester, NY: John Wiley & Sons.

Narayanan, S., Vijayakumar, P., & Govindarasu, S. (1992). Subjective assessment of sleep, fatigue, creativity and personality orientation. *Psychological Studies*, **37**, 17–25.

Neylan, T. C., Metzler, T. J., Best, S. R., Weiss, D., Fagan, J. A., Liberman, A. et al. (2002). Critical incident exposure and sleep quality in police officers. *Psychosomatic Medicine*, **64**, 345–352.

NHTSA (2002a). Wake up and get some sleep. Retrieved on May 28, 2002, from http://www.nhtsa.dot.gov/people/injury/drows_driving/index.html NHTSA (National Highway Traffic Safety Administration).

NHTSA (2002b). Sick and tired of waking up sick and tired. Retrieved on May 28, 2002, from http://www.nhtsa.dot.gov/people/injury/drows_driving/wbroch/wbrochure.pdf (National Highway Traffic Safety Administration).

Nicholson, A. N. (1987). Sleep and wakefulness of the airline pilot. *Aviation Space & Environmental Medicine*, **58**, 395–401.

NSF (1998–2002). NSF public opinion polls. Retrieved on May 28, 2002, from http://www.sleepfoundation.org/publications/execsum.html (National Sleep Foundation)

NSF (1998). Women and sleep. Retrieved on May 28, 2002, from http://www.sleepfoundation.org/publications/execsum.html (National Sleep Foundation).

Parkes, K. R. (1994). Sleep patterns, shiftwork, and individual differences: A comparison of onshore and offshore control-room operators. *Ergonomics*, **37**, 827–844.

Parkes, K. R. (1999). Shiftwork, job type, and the work environment as join predictors of health-related outcomes. *Journal of Occupational Health Psychology*, **4**, 256–268.

Parkes, K. R. (2002). Age, smoking, and negative affectivity as predictors of sleep patterns among shiftworkers in two environments. *Journal of Occupational Health Psychology*, **7**, 156–173.

Patten, C. A., Choi, W. S., Gillin, J. C., & Pierce, J. P. (2000). Depressive symptoms and cigarette smoking predict development and persistence of sleep problems in US adolescents. *Pediatrics*, **106**, E23.

Patton, D. V., Landers, D. R., & Agarwal, I. T. (2001). Legal considerations of sleep deprivation among resident physicians. *Journal of Health Law*, **34**, 377–417.

Pilcher, J. J., & Coplen, M. K. (2000). Work/rest cycles in railroad operations: Effects of shorter than 24-h shift work schedules and on-call schedules of sleep. *Ergonomics*, **43**, 573–588.

Pilcher, J. J., Ginter, D. R., & Sadowsky, B. (1997). Sleep quality versus sleep quantity: Relationships between sleep and measures of health, well-being and sleepiness in college students. *Journal of Psychosomatic Research*, **42**, 583–596.

Pilcher, J. J., & Ott, E. S. (1998). The relationships between sleep and measures of health and well-being in college students: A repeated measures approach. *Behavioral Medicine*, **23**, 170–178.

Porcu, S., Bellatreccia, A., Ferrara, M., & Casagrande, M. (1998). Sleepiness, alertness, and performance during a laboratory simulation of an acute shift of the wake-sleep cycle. *Ergonomics*, **41**, 1192–1202.

Pouliot, Z., Peters, M., Neufeld, H., & Kryger, M. H. (1997). Using self-reported questionnaire data to prioritize OSA patients for polysomnography. *Sleep*, **20**, 232–236.

Quant, J. R. (1992). The effect of sleep deprivation and sustained military operations on near visual performance. *Aviation Space & Environmental Medicine*, **63**, 172–176.

Raggatt, P. T. F. (1991). Work stress among long-distance coach drivers: A survey and correlational study. *Journal of Organizational Behavior*, **12**, 565–579.

Rau, R., Georgiades, A., Fredrikson, M., Lemne, C., & de Faire, U. (2001). Psychosocial work characteristics and perceived control in relation to cardiovascular rewind at night. *Journal of Occupational Health Psychology*, **6**, 171–181.

Rechtschaffen, A., & Bergmann, B. M. (2002). Sleep deprivation in the rat: An update of the 1998 paper. *Sleep*, **25**, 18–24.

Reyner, L. A., & Horne, J. A. (2002). Efficacy of a 'functional energy drink' in counteracting driver sleepiness. *Physiology & Behavior*, **75**, 331–335.

Rodgers, C. D., Paterson, D. H., Cunningham, D. A., Noble, E. G., Pettigrew, F. P., Myles, W. S., et al. (1995). Sleep deprivation: Effects on work capacity, self-paced walking, contractile properties and perceived exertion. *Sleep*, **18**, 30–38.

Romano, L. (2002, May 30). Towboat captain says he didn't fall asleep. *Washington Post*. Retrieved May 31, 2002, from http://www.washingtonpost.com/wp-dyn/articles/A30477-2002May29.html.

Rosa, R. R. (1991). Performance, alertness, and sleep after 3.5 yrs of 12 h shifts: A follow-up study. *Work & Stress*, **5**, 107–116.

Rosekind, M. R., Gander, P. H., Gregory, K. B., Smith, R. M., Miller, D. L., Oyung, R., et al. (1995a). Managing fatigue in operational settings: Physiological considerations and countermeasures. *Behavioral Medicine*, **21**, 157–165.

Rosekind, M. R., Gander, P. H., Gregory, K. B., Smith, R. M., Miller, D. L., Oyung, R., et al. (1995b). Managing fatigue in operational settings: An integrated approach. *Behavioral Medicine*, **21**, 166–170.

Rosenthal, L., Roehrs, T. A., & Roth, T. (1993). The Sleep–Wake Activity Inventory: A self-report measure of daytime sleepiness. *Biological Psychiatry*, **34**, 810–820.

Roth, T., Roehrs, T., & Zorick, F. J. (1982). Sleepiness: Its measurement and determinants. *Sleep*, **5**(Suppl. 2), 128–134.

Schutte, N. S., & Malouff, J. M. (1995). *Sourcebook of Adult Assessment Strategies*. New York: Plenum Press.

Scott, A. J., & LaDou, J. (1990). Shiftwork: Effects on sleep and health with recommendations for medical surveillance and screening. In A. J. Scott (ed.), *Occupational Medicine: Shiftwork* (pp. 273–300). Philadelphia: Hanley & Belfus.

Senecal, C., Vallerand, R. J., & Guay, F. (2001). Antecedents and outcomes of work–family conflict: Toward a motivational model. *Personality & Social Psychology Bulletin*, **27**, 176–186.

Seo, Y., Matsumoto, K., Park, Y., Shinkoda, H., & Noh, T. (2000). The relationship between sleep and shift system, age and chronotype in shift workers. *Biological Rhythm Research*, **31**, 559–579.

Sluiter, J. K., van der Beek, A. J., & Frings-Dresen, M. H. W. (1999). The influence of work characteristics on the need for recovery and experienced health: A study on coach drivers. *Ergonomics*, **42**, 573–583.

Smith, C. S., Folkard, S., & Fuller, J. A. (2002). Shiftwork and working hours. In J. C. Quick, & L. E. Tetrick (eds), *Handbook of Occupational Health Psychology* (pp. 163–184). Washington, DC: American Psychological Association.

Smith, C. S., Robie, C., Folkard, S., Barton, J., Macdonald, I., Smith, L., et al. (1999). A process model of shiftwork and health. *Journal of Occupational Health Psychology*, **4**, 207–218.

Smith, P. C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (ed.), *Handbook of Industrial and Organizational Psychology* (pp. 745–776). Chicago: Rand McNally.

Sparks, K., Cooper, C., Fried, Y., & Shirom, A. (1997). The effects of hours of work on health: A meta-analytic review. *Journal of Occupational and Organizational Psychology*, **70**, 391–408.

Spector, P. E., Chen, P. Y., & O'Connell, B. J. (2000). A longitudinal study of relations between job stressors and job strains while controlling for prior negative affectivity and strains. *Journal of Applied Psychology*, **85**, 211–218.

Spector, P. E., & Jex, S. M. (1998). Development of four self-report measures of job stressors and strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory, and Physical Symptoms Inventory. *Journal of Occupational Health Psychology*, **3**, 356–367.

Spelten, E., Totterdell, P., Barton, J., & Folkard, S. (1995). Effects of age and domestic commitment on the sleep and alertness of female shiftworkers. *Work & Stress*, **9**, 165–175.

Steele, M. T., Ma, O. J., Watson, W. A., & Thomas, H. A. (2000). Emergency medicine residents' shiftwork tolerance and preference. *Academic Emergency Medicine*, **7**, 670–673.

Takahashi, M., & Arito, H. (2000). Maintenance of alertness and performance by a brief nap after lunch under prior sleep deficit. *Sleep*, **23**, 813–819.

Totterdell, P., Reynolds, S., Parkinson, B., & Briner, R. B. (1994). Associations of sleep with everyday mood, minor symptoms and social interaction experience. *Sleep*, **17**, 466–475.

Totterdell, P., Spelten, E., Barton, J., Smith, L., & Folkard, S. (1995). On-shift and daily variations in self-report and performance measures in rotating-shift and permanent night nurses. *Work & Stress*, **9**, 187–197.

Van Egeren, L. F. (1992). The relationship between job strain and blood pressure at work, at home, and during sleep. *Psychosomatic Medicine*, **54**, 337–343.

Van Reeth, O., Weibel, L., Spiegel, K., Leproult, R., Dugovic, C., & Maccari, S. (2000). Interactions between stress and sleep: From basic research to clinical situations. *Sleep Medicine Reviews*, **4**, 201–219.

Vartia, M. (2001). Consequences of workplace bullying with respect to the well-being of its targets and the observers of bullying. *Scandinavian Journal of Work, Environment & Health*, **27**, 63–69.

Verlander, L. A., Benedict, J. O., & Hanson, D. P. (1999). Stress and sleep patterns of college students. *Perceptual & Motor Skills*, **88**, 893–898.

Vidacek, S., Kaliterna, L., & Radosevic-Vidacek, B. (1987). Predictive validity of individual differences measures for health problems in shiftworkers: Preliminary results. In A. Oginski, J. Pokorski, & J. Rutenfranz (eds), *Contemporary Advances in Shiftwork Research* (pp. 99–106). Krakow, Poland: Medical Academy.

Vidacek, S., Radosevicvidacek, B., Kaliterna, L., & Prizmic, Z. (1993). Individual differences in circadian-rhythm parameters and short-term tolerance to shiftwork: A follow-up study. *Ergonomics*, **36**, 117–123.
Webb, W. B., Bonnet, M., & Blume, G. (1976). A post-sleep inventory. *Perceptual and Motor Skills*, **43**, 987–993.
Wesensten, N. J., Belenky, G., Krautz, M. A., Thorne, D. R., Reichardt, R. M., & Balkin, T. J. (2002). Maintaining alertness and performance during sleep deprivation: Modafinil versus caffeine. *Psychopharmacology*, **159**, 238–247.
Wilkinson, R. T. (1960). The effect of lack of sleep on visual watch-keeping. *Quarterly Journal of Experimental Psychology*, **12**, 36–40.
Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. *Psychological Monographs,* **73**, 26.

Chapter 4

RESEARCH ON INTERNET RECRUITING AND TESTING: CURRENT STATUS AND FUTURE DIRECTIONS

Filip Lievens
Ghent University

and Michael M. Harris
University of Missouri-St Louis

INTRODUCTION

Over the last decade the Internet has had a terrific impact on modern life. One of the ways in which organizations are applying Internet technology and particularly World Wide Web (WWW) technology is as a platform for recruiting and testing applicants (Baron & Austin, 2000; Brooks, 2000; Greenberg, 1999; Harris, 1999, 2000). In fact, the use of the Internet for recruitment and testing has grown very rapidly in recent years (Cappelli, 2001). The increasing role of technology in general is also exemplified by the fact that in 2001 a technology showcase was organized for the first time during the Annual Conference of the Society for Industrial and Organizational Psychology. Recently, the American Psychological Association also endorsed a Task Force on Psychological Testing and the Internet.

It is clear that the use of Internet technology influences heavily how recruitment and testing are conducted in organizations. Hence, the emergence of Internet recruitment and Internet testing leads to a large number of research questions, many of which have key practical implications. For example, how do applicants perceive and use the Internet as a recruitment source or which Internet recruitment sources lead to more and better qualified applicants? Are Web-based tests equivalent to their paper-and-pencil counterparts? What are the effects of Internet-based testing in terms of criterion-related validity and adverse impact?

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

The rapid growth of Internet recruitment and testing illustrates that the answers to these questions have typically been taken for granted. Yet, in this chapter we aim to provide empirically based answers by reviewing the available research evidence. A second aim of our review consists of sparking future research on Internet-based recruitment and testing. Despite the fact that there exist various excellent reviews on recruitment (e.g., Barber, 1998; Breaugh & Starke, 2000; Highhouse & Hoffman, 2001) and selection (e.g., Hough & Oswald, 2000; Salgado, 1999; Schmitt & Chan, 1998) in a traditional context, no review of research on Internet-based recruitment and testing has been conducted. An exception is Bartram (2001) who primarily focused on trends and practices in Internet recruitment and testing.

This chapter has two main sections. The first section covers Internet recruitment, whereas the second one deals with Internet testing. Although we recognize that one of the implications of using the Internet is that the distinction between these two personnel management functions may become increasingly intertwined, we discuss both of them separately for reasons of clarity. In both sections, we follow the same structure. We start by enumerating common assumptions associated with Internet recruitment (testing) and by discussing possible approaches to Internet recruitment (testing). Next, we review empirical research relevant to both these domains. On the basis of this research review, the final part within each of the sections discusses recommendations for future research.

INTERNET RECRUITMENT

Assumptions Associated with Internet Recruitment

Internet recruitment has, in certain ways at least, significantly changed the way in which the entire staffing process is conducted and understood. In general, there are five common assumptions associated with Internet recruitment that underlie the use of this approach as compared with traditional methods. A first assumption is that persuading candidates to apply and accept job offers is as important as choosing between candidates. Historically, the emphasis in the recruitment model has been on accurately and legally assessing candidates' qualifications. As such, psychometrics and legal orientations have dominated the recruitment field. The emphasis in Internet recruitment is on attracting candidates. As a result, a marketing orientation has characterized this field.

A second assumption is that the use of the Internet makes it far easier and quicker for candidates to apply for a job. In years past, job-searching was a more time-consuming activity. A candidate who wished to apply for a job would need to locate a suitable job opportunity, which often involved searching through a newspaper or contacting acquaintances. After locating potentially suitable openings, the candidate would typically have to prepare a cover

letter, produce a copy of his or her resume, and mail the package with the appropriate postage. By way of comparison, the Internet permits a candidate to immediately seek out and search through thousands of job openings. Application may simply involve sending a resume via email. In that way, one can easily and quickly apply for many more jobs in a far shorter period of time than was possible before Internet recruitment was popularized. In fact, as discussed later, an individual perusing the Internet may be drawn quite accidentally to a job opening.

Third, one typically assumes that important information about an organization may be obtained through the Internet. The use of the Internet allows organizations to pass far more information in a much more dynamic and consistent fashion to candidates than was the case in the past. Candidates may therefore have much more information at their disposal before they even decide to apply for a job than in years past. In addition, candidates can easily and quickly search for independent information about an organization from diverse sources, such as chatrooms, libraries, and so forth. Thus, unlike years past where a candidate may have applied for a job based on practically no information, today's candidate may have reviewed a substantial amount of information about the organization before choosing to apply.

A fourth assumption is that applicants can be induced to return to a web site. A fundamental concept in the use of the Internet is that web sites can be designed to attract and retain user interest. Various procedures have been developed to retain customer interest in a web site, such as cookies that enable the web site to immediately recall a customer's preferences. Effective Internet recruitment programs will encourage applicants to apply and return to the web site each time they search for a new job. A final assumption refers to cost issues, namely that Internet recruitment is far less expensive than traditional approaches. Although the cost ratio is likely to differ from situation to situation, and Internet recruitment and traditional recruitment are not monolithic approaches, a reasonable estimate is that Internet recruitment is one-tenth of the cost of traditional methods and the amount of time between recruitment and selection may be reduced by as much as 25% (Cober, Brown, Blumental, Doverspike, & Levy, 2000).

Approaches to Internet Recruitment

We may define Internet recruitment as any method of attracting applicants to apply for a job that relies heavily on the Internet. However, it should be clear that Internet recruitment is somewhat of a misnomer because there are a number of different approaches to Internet recruitment. The following describes five important Internet recruitment approaches. We start with some older approaches and gradually move to more recent ones. This list is neither meant to be exhaustive nor comprehensive as different approaches to Internet recruitment are evolving regularly.

Company web sites

Company web sites represent one of the first Internet-based approaches to recruiting. Many of these web sites also provide useful information about the organization, as well as a mechanism for applying for these jobs. A study in 2001 by iLogos showed that of the Global 500 companies, 88% had a company Internet recruitment site, reflecting a major surge from 1998, when only 29% of these companies had such a web site. Almost all North American Global 500 companies (93%) have a company Internet recruitment site. Most applicants would consider a medium- to large-size company without a recruitment web site to be somewhat strange; indeed, one report indicated that of 62,000 hires at nine large companies, 16% were initiated at the company Internet recruitment site (Maher & Silverman, 2002). Given these numbers, and the relatively low cost, it would seem foolish for an organization not to have a company Internet recruitment site.

Job boards

Another early approach to Internet-based recruiting was the job board. Monster Board (www.monster.com) was one of the most successful examples of this approach. Basically, the job board is much like a newspaper listing of job opportunities, along with resumes of job applicants. The job board's greatest strength is the sheer numbers of job applicants listing resumes; it has been estimated that they contain 5 million unique resumes (Gutmacher, 2000). In addition, they enable recruiters to operate 24 hours a day, examine candidates from around the world, and are generally quite inexpensive (Boehle, 2000). A major advantage of the job board approach for organizations is that many people post resumes and that most job boards provide a search mechanism so that recruiters can search for applicants with the relevant skills and experience. A second advantage is that an organization can provide extensive information, as well as a link to the company's web site for further information on the job and organization.

The extraordinary number of resumes to be found on the web, however, is also its greatest weaknesses; there are many recruiters and companies competing for the same candidates with the same access to the job boards. Thus, just as companies have the potential to view many more candidates in a short period of time, candidates have the opportunity to apply to many more companies. Another disadvantage is that having access to large numbers of candidates means that there are potentially many more applicants that have to be reviewed. Finally, many unqualified applicants may submit resumes, which increases the administrative time and expense. As an example, Maher and Silverman (2002) reported one headhunter who posted a job ad for an engineering vice president on five job boards near the end of the day. The next morning, he had over 300 emailed resumes, with applicants ranging from

chief operating officers to help-desk experts. Despite the amount of attention and use of job boards, relatively few jobs may actually be initiated this way; combined together, the top four job boards produced only about 2% of actual jobs for job hunters (Maher & Silverman, 2002).

e-Recruiting

A completely different approach to Internet-based recruiting focuses on the recruiter searching on-line for job candidates (Gutmacher, 2000). Sometimes referred to as a 'meta-crawler' approach (Harris & DeWar, 2001), this approach emphasizes finding the 'passive' candidate. In addition to combing through various chat rooms, there are a number of different techniques that e-recruiters use to ferret out potential job candidates. For example, in a technique called 'flipping', recruiters use a search engine, like Altavista.com, to search the WWW for resumes with links to a particular company's web site. Doing so may reveal the resumes, email addresses, and background information for employees associated with that web site. Using a technique known as 'peeling', e-recruiters may enter a corporate web site and 'peel' it back, to locate lists of employees (Silverman, 2000).

The major advantage of this technique is the potential to find outstanding passive candidates. In addition, because the e-recruiter chooses whom to approach, there will be far fewer candidates and especially far fewer unqualified candidates generated. There are probably two disadvantages to this approach. First, because at least 50,000 people have been trained in these techniques, and companies have placed firewalls and various other strategies in place to prevent such tactics, the effectiveness of this technique is likely to decline over time (Harris & DeWar, 2001). Second, some of these techniques may constitute hacking, which at a minimum may be unethical and possibly could be a violation of the law.

Relationship recruiting

A potentially major innovation in Internet recruitment is called relationship recruiting (Harris & DeWar, 2001). A major goal of relationship recruiting is to develop a long-term relationship with 'passive' candidates, so that when they decide to enter the job market, they will turn to the companies and organizations with which they have developed a long-term relationship (Boehle, 2000). Relationship recruitment relies on Internet tools to learn more about web-visitors' interests and experience and then email regular updates about careers and their fields of interest. When suitable job opportunities arise, an email may be sent to them regarding the opportunity. For an interesting example, see http://www.futurestep.com. Probably the major advantage of this approach is that passive applicants may be attracted to jobs with a good fit. Over time, a relationship of trust may develop that will produce candidates who return to the web site whenever they are seeking

jobs, thus creating a long-term relationship. At this point, it is unclear what disadvantages, if any, there are to relationship recruitment. One possibility may be that relationship recruitment may simply fail to generate enough applicants for certain positions.

Surreptitious approaches

Perhaps the most recent approach to Internet recruitment is the surreptitious or indirect approach. The best example is provided by www.salary.com, which provides free salary survey information. Because the web site enables one to request information by job title and geographic location, information about potential job opportunities can be automatically displayed. This site provides additional services (e.g., a business card, which can be sent with one's email address, to potential recruiters) that facilitate recruitment efforts. We imagine that if it is not happening yet, 'pop-up' ads for jobs may soon find their way to the Internet. Although it is too early to assess the strengths and weaknesses of surreptitious approaches to recruitment, they would appear to be a potentially useful way to attract passive job applicants. On the other hand, some of these techniques may be perceived as being rather offensive and overly direct.

Previous Research

Despite the rapid emergence of Internet recruitment approaches, research studies on Internet recruitment are very sparse. To the best of our knowledge, the only topic that has received some empirical research attention is how people react to various Internet-based recruitment approaches.

Weiss and Barbeite (2001) focused on reactions to Internet-based job sites. To this end, they developed a web-based survey that addressed the importance of job site features, privacy issues, and demographics. They found that the Internet was clearly preferred as a source of finding jobs. In particular, respondents liked job sites that had few features and required little personal information. Yet, older workers and women felt less comfortable disclosing personal information at job sites. Men and women did not differ in terms of preference for web site features, but women were less comfortable providing information online. An experimental study by Zusman and Landis (2002), who compared potential applicants' preferences for web-based versus traditional job postings, did not confirm the preference for web-based job information. Undergraduate students preferred jobs on traditional paper-and-ink materials over web-based job postings. Zusman and Landis also examined the extent to which the quality of an organization's web site attracted applicants. In this study, poor-quality web sites were defined as those using few colors, no pictures, and simple fonts, whereas high-quality web sites were seen as the opposite. Logically, students preferred jobs on high quality web

pages to those on lower quality pages. Scheu, Ryan, and Nona (1999) confirmed the role of web site aesthetics. In this study, impressions of a company's web site design were positively related to intentions to apply to that company. It was also found that applicant perceptions of a company changed after visiting that company's web site.

Rozelle and Landis (2002) gathered reactions of 223 undergraduate students to the Internet as a recruitment source and more traditional sources (i.e., personal referral, college visit, brochure about university, video about university, magazine advertisement). On the basis of the extant recruitment source literature (see Zottoli & Wanous, 2000, for a recent review), they classified the Internet as a more formal source. Therefore, they expected that the Internet would be perceived to be less realistic, leading to less positive post-selection outcomes (i.e., less satisfaction with the university). Yet, they found that the Internet was seen as more realistic than the other sources. In addition, use of the university web page as a source of recruitment information was not negatively correlated with satisfaction with the university. According to Rozelle and Landis, a possible explanation for these results is that Internet recruitment pages are seen as less formal recruitment sources than, for example, a brochure because of their interactivity and flexibility.

Whereas the previous studies focused on web-based job postings, it is also possible to use the Internet to go one step further and to provide potential applicants with realistic job previews (Travagline & Frei, 2001). This is because Internet-based, realistic job previews can present information in a written, video, or auditory format. Highhouse, Stanton, and Reeve (forthcoming) examined reactions to such Internet-based realistic job previews (e.g., the company was presented with audio and video excerpts). Interestingly, Highhouse et al. did also not examine retrospective reactions. Instead, they used a sophisticated micro-analytic approach to examine on-line (i.e., instantaneous) reactions to positive and negative company recruitment information. Results showed that positive and negative company information in a web-based job fair elicited asymmetrically extreme reactions such that the intensity of reactions to positive information were greater than the intensity of reactions to negative information on the same attribute.

Dineen, Ash, and Noe (2002) examined another aspect of web-based recruitment, namely the possibility to provide tailored on-line feedback to candidates. In this experimental study, students were asked to visit the career web page of a fictitious company that provided them with information about the values of the organization and with an interactive 'fit check' tool. In particular, participants were told whether they were a 'high' or a 'low' fit with the company upon completion of a web-based person–organization fit inventory. Participants receiving feedback that indicated high P–O fit were significantly more attracted to the company than participants receiving no feedback. Similarly, participants receiving low-fit feedback were significantly less attracted than those receiving no feedback.

Finally, Elgin and Clapham (2000) did not investigate applicant reactions to Internet-based recruitment but concentrated on the reactions of recruiters. The central research question was whether recruiters associated different attributes with job applicants with an electronic resume vs. job applicants with paper resumes. Results revealed that the electronic resume applicant was perceived as possessing better overall qualifications than the applicants using paper resumes. More detailed analyses further showed that the paper resume applicant was perceived as more friendly, whereas the electronic resume applicant was viewed as significantly more intelligent and technologically advanced.

Although it is difficult to draw firm conclusions due to the scarcity of research, studies generally yield positive results for the Internet as a recruitment mechanism. In fact, applicants seem to react favorably to Internet job sites and seem to prefer company web pages over more formal recruitment sources. There is also initial evidence supporting other aspects of Internet-based recruitment such as the possibility of offering realistic job previews and online feedback.

Recommendations for Future Research

Because of the apparent scarcity of research on Internet-based recruitment, this subsection discusses several promising routes for future research, namely applicant decision processes in Internet recruitment (i.e., decisions regarding which information to use and how to use that information), the role that the Internet plays in recruitment, and the effects of Internet recruitment on the turnover process.

How do applicants decide which sources to use?

Although there is a relatively large literature concerning applicant source (e.g., newspaper, employee referral) and applicant characteristics in the broader recruitment literature (Barber, 1998; Zottoli & Wanous, 2000), there is practically no research on how applicants perceive different Internet sources. In other words, do applicants perceive that some Internet sources of jobs are more useful than others? Several factors may play a role here. One factor, not surprisingly, would be the amount of available information and the quality of the jobs. A second factor may be the degree to which confidentiality and privacy is perceived to exist (for more information and discussion of this topic, see the section, 'Draw on psychological theories to examine Internet-based testing applications', about privacy in the section on Internet-based testing; see also the aforementioned study of Weiss & Barbeite, 2001). A third factor may be aesthetic qualities, such as the attractiveness of the graphics (see Scheu et al., 1999; Zusman & Landis, 2002).

Technical considerations, such as the quality of the search engines, the speed with which the web site operates, and related issues (e.g., frequency of crashes), comprise the fourth factor.

Cober, Brown, Blumental, and Levy (2001) presented a three-stage model of the Internet recruitment process. The first stage in the model focuses on persuading Internet users to review job opportunities on the recruitment site. The model assumes that at this stage in the process, applicants are primarily influenced by the aesthetic and affective appeal of the web site. The second stage of the process focuses on engaging applicants and persuading them to examine information. This stage in turn comprises three substages: fostering interest, satisfying information requirements, and building a relationship. At this stage, applicants are primarily swayed by concrete information about the job and company. The final stage in this model is the application process, wherein people decide to apply on-line for a position. Cober et al. (2001) rated a select group of companies' recruitment web sites on characteristics such as graphics, layout, key information (e.g., compensation), and reading level. Using this coding scheme, they reported that most of these companies had at least some information on benefits and organizational culture. Relatively few of these companies provided information about such items as vision or future of the organization. The estimated reading level was at the 11th-grade level. Interestingly, reading level was negatively correlated with overall evaluation of the company's recruitment web site. The more aesthetically pleasing the web site, the more positively it was rated as well. Given the typology developed by Cober et al., the next logical step would be to study the effect on key measures such as number of applicants generated, how much time was spent viewing the web site, and the number of job offers accepted.

We believe that certain factors may moderate the importance of the things that we have already mentioned. One moderator may be the reputation of the organization; individuals may focus more on one set of factors when considering an application to a well-regarded organization than when viewing the site of an unknown organization. We also suspect that factors that initially attract job-seekers may be different than the factors that encourage candidates to return to a web site. Specifically, while aesthetic and technical factors may initially affect job-seekers, they are likely to play a less prominent role as job-seekers gain experience in applying for jobs. Cober et al.'s (2001) model and typology appears to be a good way to begin studying applicant decision processes.

Resource exchange theory (Brinberg & Ganesan, 1993; Foa, Converse, Tornblom, & Foa, 1993) is another model that may be helpful in understanding the appeal of different Internet-based recruitment sites. Yet, to our knowledge, this theory has not been extensively applied in the field of industrial and organizational (I/O) psychology. Briefly stated, resource exchange theory assumes that all resources (e.g., physical, psychological, etc.)

can be sorted into six categories: information, money, goods, services, love, and status. Moreover, these six categories can be classified along two dimensions: particularism and concreteness. Particularism refers to the degree to which the source makes a difference—love is very high on particularism because it is closely tied to a specific source (i.e., person), while money is very low on particularism because it is the same, no matter what the source. Services, on the other hand, are higher on particularism than goods. The second dimension, concreteness, refers to the degree to which the resource is symbolic (e.g., status) or tangible (e.g., goods). Not surprisingly, status and information are the most symbolic, while goods and services are the most concrete (see Foa et al., 1993, for a good background to this theory).

Beyond the classification aspect of the theory, there are numerous implications. For present purposes, we will focus on some of the findings of Brinberg and Ganesan (1993), who applied this theory to product positioning, which we believe is potentially relevant to understanding job-seeker use of Internet recruitment. Specifically, Brinberg and Ganesan examined whether the category in which a consumer places a specific resource can be manipulated. For example, jewelry, described to a subject as a way to show someone that he or she cares, was more likely to be classified as being in the 'love' category than was jewelry, described to a subject as serving many practical purposes for an individual, which was more likely to be classified as being in the 'service' category. Based on the assumption that the perceived meaning of a particular product, in this case an Internet recruitment site, affects the likelihood of purchase (in this case, joining or participating), resource exchange theory may provide some interesting predictions. For example, by selling an Internet recruitment site as a service (which is more particularistic and more concrete) rather than information, job-seekers may be more likely to join. Thus, we would predict that the greater the match between what job-seekers are looking for in an Internet site (e.g., status and service) and the image that the Internet site offers, the more likely job-seekers will use the Internet site.

Finally, the elaboration likelihood model (Larsen & Phillips, 2001; Petty & Cacioppo, 1986) may be fruitfully used to understand how applicants choose Internet recruitment sites. Very briefly, the elaboration likelihood model separates variables into central cues (e.g., information about pay) and peripheral cues (e.g., aesthetics of the web site). Applicants must be both able and motivated to centrally process the relevant cues. When they are either not motivated or not able to process the information, they will rely on peripheral processing and utilize peripheral cues to a larger extent. Furthermore, decisions made using peripheral processing are more fleeting and likely to change than decisions made using central processing. We would expect that aesthetic characteristics are peripheral cues and that their effect is often fleeting. In addition, we would expect that first-time job-seekers use peripheral processing more frequently than do veteran job-seekers. Clever research designs using Internet sites should be able to test some of these assertions.

How is Internet-based information used by applicants?

As described above, one key assumption of Internet recruiting is that important information about an organization can be easily and quickly obtained through the use of search engines, as well as company-supplied information. A number of interesting research questions emanate from this assumption. First, there are many different types of web sites that may contain information about an organization. We divide these into three types: official company web sites, news media (e.g., www.lexisnexis.com), and electronic bulletin boards (e.g., www.vault.com). Paralleling earlier research in the recruitment area (Fisher, Ilgen, & Hoyer, 1979), it would be interesting to determine how credible each of these sources is perceived to be. For example, information from a chatroom regarding salaries at a particular organization may be considered more reliable than information about salaries offered in a company-sponsored web site. Likewise, does the source credibility depend upon the facet being considered? For example, is information regarding benefits considered more credible when it comes from official company sources, while information about the quality of supervision is perceived to be more reliable when coming from a chatroom?

A related question of interest is what sources of information candidates actually do use at different stages in the job search process. Perhaps certain sources are more likely to be tapped than others early in the recruitment process, whereas different sources are likely to be scrutinized later in the recruitment process. It seems likely, for example, that information found on the company web site may be weighted more heavily in the early part of the recruitment process (e.g., in the decision to apply) than in later stages of the process (e.g., in choosing between different job offers). In later stages of the recruitment process, particularly when a candidate is choosing between competing offers, perhaps electronic bulletin boards are more heavily weighted. Longitudinal research designs, which have already been used in the traditional recruitment domain (e.g., Barber, Daly, Giannantonio, & Phillips, 1994; Saks & Ashforth, 2000), should be used to address these questions.

Finally, researchers should explore the use of Internet-based information vs. other sources of information about the organization (see Rozelle & Landis, 2002). Besides the Internet, information may be obtained from a site visit of the organization, where candidates speak with their future supervisor, co-workers, and possibly with subordinates. As already noted above, there exists a voluminous literature on information sources in recruitment. How information from those traditional sources is integrated with information obtained from the Internet should be studied more carefully, particularly when contradictory information is obtained from multiple sources. Again, longitudinal research using realistic fields settings is needed here.

What role does the Internet play in recruitment?

Given the number of resumes on-line and use of Internet recruitment sites, one may conclude that the Internet plays a major role in recruitment. Yet, surveys indicate that networking is still by far the most common way to locate a job. There are several questions that should be investigated regarding the role of the Internet in recruitment. First, how are job-seekers using the Internet—is the Internet their first strategy in job search? Is it supplanting other methods, such as networking? Second, it seems likely that a host of demographic variables will affect applicant use of the Internet versus other recruitment methods. Sharf (2000) observed that there are significant differences in the percentage of households possessing Internet access, depending on race, presence of a disability, and income. Organizations may find that heavy dependence on Internet recruitment techniques hampers their efforts in promoting workforce diversity (Stanton, 1999). Finally, more research should be performed comparing the different methods of Internet recruitment. For instance, e-recruiting should be compared with traditional headhunting methods. We suspect that applicants may prefer e-recruiting over face-to-face or even telephone-based approaches. First, email is perceived to be more private and anonymous in many ways as compared with the telephone. Second, unlike the telephone, email allows for an exchange of information even when the sender and recipient are not available at the same time. Whether or not different Internet recruitment methods have different effects on applicants remains to be studied.

The effects of Internet recruitment on the turnover process

To date, there has been little discussion about the impact of Internet recruitment on the turnover process. However, we believe that there are various areas where the use of Internet recruitment may affect applicants' decision to leave their present organization, including the decision to quit, the relationships between withdrawal cognitions, job search, and quitting, and the costs of job search.

With regard to the decision to quit, there has been a plethora of research. The most sophisticated models of the turnover process include job search in the sequence of events (Hom & Griffeth, 1995). One of the most recent theories, known as the unfolding model (Lee & Mitchell, 1994), posits that the decision by an employee to leave his or her present organization is based on one of four 'decision paths'. Which of the four 'decision paths' is chosen depends on the precipitating event that occurs. In three of the decision paths, the question of turnover is raised when a shock occurs. A shock is defined as 'a specific event that jars the employee to make deliberate judgments about his or her job' (Hom and Griffeth, 1995, p. 83). When one's company is acquired by another firm, for example, this may create a shock to an em-

ployee, requiring the employee to think more deliberately about his or her job. According to Mitchell, Holtom, and Lee (2002), Path-3 leavers often initiate the turnover process when they receive an unsolicited job offer. It seems plausible, then, that with the frequency of individuals using Internet recruitment, there will be a significant increase in the number of individuals using the third decision path. As explained by Mitchell et al., individuals using the third path are leaving for a superior job. Thus, individuals who read Internet job postings may realize that there are better job alternatives, which to use Lee and Mitchell's terminology, prompts them to review the decision to remain with their current employer. Research is needed to further understand the use of Internet recruitment and turnover processes, using the unfolding model. Are there, for example, certain Internet recruitment approaches (e.g., e-recruiting) that are particularly likely to induce Path-3 processes? What type of information should these approaches use to facilitate turnover?

A second area relates to the relationships between withdrawal cognitions, job search, and quitting. As discussed by Hom and Griffeth (1995), one of the debates in the turnover literature concerns the causal paths among withdrawal cognitions, job search, and quitting. Specifically, there have been different opinions as to whether employees decide to quit and then go searching for alternative jobs, or whether employees first go searching for alternative jobs and then decide to quit their present company. Based on the existing evidence, Hom and Griffeth argue for the former causal ordering. However, using the assumption of Lee and Mitchell that different models of turnover may be relevant for different employees, it seems plausible that individuals using Internet recruitment might follow the latter causal order. In order words, individuals reviewing Internet job postings 'just for fun' may locate opportunities of interest, which compare more favorably than their current position. The existence of Internet recruitment may therefore affect the relationship between withdrawal cognitions, job search, and quitting.

The costs of job search constitute a third possible area where the use of Internet recruitment may affect applicants' decisions to leave their present organization. As we noted above, the use of the Internet may greatly reduce the cost of job searching. Although there has been little research done on job search activity by I/O psychologists, it seems reasonable that the expectancy model, which includes an evaluation of the costs and benefits and the likelihood of success, will determine the likelihood of one engaging in job search behavior. Given that the use of Internet recruitment can greatly reduce the costs to a job-searcher, it seems reasonable to assume that individuals will be more likely to engage in a job search on a regular basis than in the past. Models of the turnover process in general, and the job search process in specific, should consider the perceived costs versus benefits of job hunting for the employee. In all likelihood, as the costs decline, employees would be more likely to engage in job hunting.

In sum, there are some interesting possible effects of Internet recruitment on turnover processes. In general, research linking recruitment theories and turnover theories appears to be lacking. It is time to integrate these two streams of research.

INTERNET TESTING

Common Advantages Associated with Internet Testing

The use of the Internet is not only attractive for recruitment purposes. There are also a number of factors that lead organizations to invest in the web for testing purposes. On the one hand, testing candidates through the Internet builds further on the advantages inherent in computerized testing. Similar to computerized testing (McBride, 1998), Internet testing involves considerable test administration and scoring efficiencies because test content can be easily modified, paper copies are no longer needed, test answers can be captured in electronic form, errors can be routinely checked, tests can be automatically scored, and instant feedback can be provided to applicants. This administrative ease may result in potentially large savings in costs and turnaround time, which may be particularly important in light of tight labor markets. Akin to computerized testing, Internet-based testing also enables organizations to present items in different formats and to measure other aspects of applicant behavior. In particular, items might be presented in audio and video format, applicants' response latencies might be measured, and items might be tailored to the latent ability of the respondents.

On the other hand, web-based testing also has various additional advantages over computerized testing (Baron & Austin, 2000; Brooks, 2000). In fact, the use of the web for presenting test items and capturing test-takers' responses facilitates consistent test administration across many divisions/sites of a company. Further, because tests can be administered over the Internet, neither the employer nor the applicants have to be present in the same location, resulting in increased flexibility for both parties. Hence, given the widespread use of information technology and the globalization of the economy, Internet-based testing might expand organizations' access to other and more geographically diverse applicant pools.

Approaches to Internet Testing

Because of the rapid growth of Internet testing and the wide variety of applications, there are many ways to define Internet-based testing. A possible straightforward definition is that it concerns the use of the Internet or an intranet (an organization's private network) for administering tests and inventories in the context of assessment and selection. Although this definition (and this chapter) focus only on Internet-based tests and Internet-based

inventories, it is also possible to use the Internet (through videoconference) for conducting employment interviews (see Straus, Miles, & Levesque, 2001).

The wide variety in Internet-based testing applications is illustrated by looking at two divergent examples of current Internet-based testing applications. We chose these two examples for illustration purposes because they represent relative extremes. First, Baron and Austin (2000) developed a web-based cognitive ability test. This test was a timed, numerical reasoning test with business-related items and was used after an on-line application and before participation in an assessment center. Applicants could fill in the test whenever and wherever they wanted to. There was no test administrator present. The test was developed according to item response theory principles so that each applicant received different items tailored to his/her ability. In addition, there existed various formats (e.g., text, table, or graphic) for presenting the same item content so that it was highly improbable that candidates received the same items. Baron and Austin (2000) also built other characteristics into the numerical reasoning test to counter user identification problems and possible breaches to test security. For example, the second part of the test was administered later in the selection process in a supervised context so that the results of the two sessions could be compared. In addition, applicants were required to fill in an honesty contract, which certified that they and nobody else completed the Web-based test. The system also allowed candidates to take the test only once and encrypted candidate responses for scoring and reporting.

Second, Greenberg (1999) presented a radically different application of Internet testing. Probably, this application is more common in nowadays organizations. Here applicants were not allowed to log on where and when they wanted to. Instead, applicants were required to log on to a web site from a standardized and controlled setting (e.g., a company's test center). A test administrator supervised the applicants. Hence, applicants completed the tests in structured test administration conditions.

Closer inspection of these examples and other existing web-based testing applications illustrates (e.g., Coffee, Pearce, & Nishimura, 1999; Smith, Rogg, & Collins, 2001) that web-based testing can vary across several categories/dimensions. We believe that at least the following four dimensions should be distinguished: (1) the purpose of testing, (2) the selection stage, (3) the type of test, and (4) the test administration conditions. Although these four dimensions are certainly not orthogonal, we discuss each of them separately.

Regarding the first dimension of *test purpose*, Internet testing applications are typically divided into applications for career assessment purposes vs. applications for hiring purposes. At this moment, tests for career assessment purposes abound on the Internet (see Lent, 2001; Oliver & Whiston, 2000, for reviews). These tests are often provided for free to the general public,

although little is known about their psychometric properties. The other side of the continuum consists of organizations that use tests for hiring purposes. Given this consequentiality, it is expected that these tests adhere to professional standards (Standards for Educational and Psychological Testing, 1999) so that they have adequate psychometric properties.

A second question deals with the *stage in the selection process* wherein organizations are using Internet testing. For example, some organizations might use Internet-based testing applications for screening ('selecting out') a large number of applicants and for reducing the applicant pool to more manageable proportions. Conversely, other organizations might use Internet-based testing applications at the final stage of the selection process to 'select in' already promising candidates.

A third dimension pertains to the *type of test* administered through the WWW. In line with the computerized testing literature, a relevant distinction opposes cognitive-oriented measures vs. noncognitive-oriented measures. Similarly, one can make a distinction between tests with a correct answer (e.g., cognitive ability tests, job knowledge tests, situational judgment tests) vs. tests without a correct answer (e.g., personality inventories, vocational interest inventories). At this moment, organizations most frequently seem to use noncognitive-oriented web-based measures. In fact, Stanton and Rogelberg (2001a) conducted a small survey of current web-based hiring practices and concluded that virtually no organizations are currently using the Internet for administering cognitive ability tests.

The fourth and last dimension refers to the *test administration conditions* and especially to the level of control and standardization by organizations over these conditions. Probably, this dimension is the most important because it is closely related to the reliability and validity of psychological testing (Standards for Educational and Psychological Testing, 1999). In Internet testing applications, test administration conditions refer to various aspects such as the time of test administration, the location, the presence of a test administrator, the interface used, and the technology used. Whereas in traditional testing, the control over these aspects is typically in the hands of the organizations, this is not necessarily the case in Internet testing applications. For example, regarding the *time* of test administration, some organizations enable applicants to log on whenever they want to complete the tests (see the example of Baron & Austin, 2000). Hence, they provide applicants with considerable latitude. Other organizations decide to exert a lot of control. In this case, organizations provide applicants access to the Internet test site only at fixed, predetermined times.

Besides test administration time, test administration *location* can also vary in web-based testing applications. There are organizations that allow applicants to log on where they want. For example, some applicants may log on to the web site from their home, others from their office, and still others from a computing room. Some people may submit information in a noisy computer

lab, whereas others may be in a quiet room (Buchanan & Smith, 1999a; Davis, 1999). This flexibility and convenience sharply contrast to the standardized location (test room) in other web-based testing applications. Here, applicants either go to the company's centralized test center or to the company's multiple geographically dispersed test centers or supervised kiosks.

Another aspect of test administration conditions of Internet testing applications refers to the decision as to whether or not a *test-administrator* is used. This dimension of web-based testing is also known as proctored (supervised) vs. unproctored (unsupervised) web-based testing. In some cases, there is no test-administrator to supervise applicants. When no test-administrator is present, organizations lack control over who is conducting the test. In addition, there is no guarantee that people do not cheat by using help from others or reference material (Baron & Austin, 2000; Greenberg, 1999; Stanton, 1999). Therefore, in most Internet-based testing applications, a test-administrator is present to instruct testees and to ensure that they do not use dishonest means to improve their test performance (especially on cognitive-oriented measures).

In web-enabled testing, test administration conditions also comprise the type of *user interface* that organizations use (Newhagen & Rafaeli, 2000). Again, the type of interface used may vary to a great extent across Internet-based testing applications. At one side of the continuum, there are Internet-based testing applications that contain a very restrictive user interface. For example, some organizations decide to increase standardization and control by heavily restricting possible applicant responses such as copying or printing the items for test security reasons. Other restrictions consist of requirements asking applicants (a) to complete the test within a specific time limit, (b) to complete the test in one session, (c) to fill in all necessary information on a specific test form prior to continuing, and (d) neither to skip nor backtrack items. When applicants do one of these things, a warning message is usually displayed. At the other side of the continuum, some organizations decide to give applicants more latitude in completing Internet-based tests.

Finally, the *WWW technology* is also an aspect of test administration conditions that may vary substantially across Internet-based testing applications. In this context, we primarily focus on how technology is related to test administration conditions (see Mead, 2001, for a more general typology of WWW technological factors). Some organizations invest in technology to exert more control and standardization over test administration. For example, to guarantee to applicants that the data provided are 'secure' (are not intercepted by others), organizations may decide to use encryption technology. In addition, organizations may invest in computer and network resources to assure the speed and reliability of the Internet connection. To ensure that the person completing the test is the applicant, in the near future

organizations may decide to use web cams, typing patterns, fingerprint scanning, or retinal scanning (Stanton & Rogelberg, 2001a). All these technological interventions are especially relevant when organizations have no control over other aspects of the web-based test-administration (e.g., absence of a test-administrator). Other organizations may decide not to invest in these new technologies. Instead, they may invest in a proctored test environment (e.g., use of a test administrator to supervise applicants).

Taken together, these examples and this categorization of Internet testing highlight that Internet-based testing applications may vary considerably. Unfortunately, no data have been gathered about the frequency of use of the various forms of Web-based testing in consultancy firms and companies. In any case, all of this clearly shows that there is no 'one' way of testing applicants through the Internet and that Web-based testing should not be regarded as a monolithic entity. Hence, echoing what we have said about Internet recruitment, we believe that the terms 'Internet testing' or 'Web-based testing' are misnomers and should be replaced by 'Internet testing applications' or 'Web-based testing applications'.

Previous Research

Although research on Internet testing is lagging behind Internet testing practice, the gap is less striking than for Internet recruitment research. This is because empirical research on Internet testing has proliferated in recent years. Again, most of the studies that we retrieved were in the conference presentation format and had not been published yet. Note also that only a limited number of research topics have been addressed. The most striking examples are that, to the best of our knowledge, neither the criterion-related validity of Internet testing applications nor the possible adverse impact of Internet testing applications have been put to scrutiny. Moreover, most studies have treated Internet testing as a monolithic entity, ignoring the multiple dimensions of Internet testing discussed above. The remainder of this section summarizes the existing studies under the following two headings: measurement equivalence and applicant perceptions.

Measurement equivalence

In recent years, a sizable amount of studies have examined whether data collected through the WWW are similar to data collected via the traditional paper-and-pencil format. Three streams of research can be distinguished. A first group of studies investigated whether Internet data collection was different from 'traditional' data collection (see Stanton & Rogelberg, 2001b and Simsek & Veiga, 2001, for excellent reviews). Strictly speaking, this first group of studies dealt not really with Internet testing application because most of them were not conducted in a selection context. Instead, they focused

on data collection of psychosocial data (Buchanan & Smith, 1999a, b; Davis, 1999; Joinson, 1999; Pasveer & Ellard, 1998; Pettit, 1999), survey data (Burnkrant & Taylor, 2001; Hezlett, 2000; Magnan, Lundby, & Fenlason, 2000; Spera & Moye, 2001; Stanton, 1998), or multisource feedback data (Fenlason, 2000). In general, no differences or minimal differences between Internet-based data collection and traditional (paper-and-pencil) data collection were found.

A second group of studies did focus on selection instruments. Specifically, these studies examined the equivalence of selection instruments administered in either Web-based vs. traditional contexts. Mead and Coussons-Read (2002) used a within-subjects design to assess the equivalence of the Sixteen Personality Factor Questionnaire. Sixty-four students were recruited from classes and completed first the paper-and-pencil version and about two weeks later the Internet version. Cross-mode correlations ranged between 0.74 to 0.93 with a mean of 0.85, indicating relatively strong support for equivalence. Although this result is promising, a limitation is that the study was conducted with university students. Two other studies examined similar issues with actual applicants. Reynolds, Sinar, and McClough (2000) examined the equivalence of a biodata-type instrument among 10,000 actual candidates who applied for an entry-level sales position. Similar to Mead and Coussons-Read (2002), congruence coefficients among the various groups were very high. However, another study (Ployhart, Weekley, Holtz, & Kemp, 2002) reported somewhat less positive results with a large group of actual applicants for a teleservice job. Ployhart et al. used a more powerful procedure such as multiple group, confirmatory factor analysis to compare whether an Internet-based administration of a Big Five-type personality inventory made a difference. Results showed that the means on the Web-based personality inventory were lower than the means on the paper-and-pencil version. Although the factor structures took the same form in each administration condition, the factor structures were partially invariant, indicating that factor loadings were not equal across administration formats.

Finally, a third set of studies concentrated on the equivalence of different approaches to Internet testing. Oswald, Carr, and Schmidt (2001) manipulated not only test administration format but also test administration setting to determine their effects on measurement equivalence. In their study, 410 undergraduate students completed ability tests (verbal analogies and arithmetic reasoning) and a Big Five personality inventory (a) either in paper-and-pencil or Internet-based format and (b) either in supervised or unsupervised testing settings. Oswald et al. (2001) hypothesized that ability and personality tests would be less reliable and have a less clear factor structure under unsupervised and therefore less standardized conditions. Preliminary findings of multiple group confirmatory factor analyses showed that for the personality measures administered in supervised conditions, model fit tended to support measurement invariance. Conversely,

unsupervised measures of personality tended not to show good fit, lending support to the original hypothesis. Remarkably, for cognitive ability measures, both supervised and unsupervised conditions had a good fit. In another study, Beaty, Fallon, and Shepherd (2002) used a within-subjects design to examine the equivalence of proctored (supervised) vs. unproctored (unsupervised) Internet testing conditions. So, interestingly, these authors did not also treat Internet-based testing as a monolithic entity. Another interesting aspect of the study was that real applicants were used. First, applicants completed the unproctored test at home or at work. Beaty et al. found that the average score of the applicants was 35.3 (SD = 6.5). Next, the best 76 candidates were invited to complete a parallel form of the test in a proctored test session. The average score for these candidates in the proctored testing session was 42.2 (SD = 2.0). In comparison, this same group had an average test score of 44.1 (SD = 4.9) in the unproctored test session ($t = 3.76$, $p < 0.05$). Although significant, the increase in test scores in unsupervised Web-based testing environments (due to cheating such as having other people fill in the test) seems to be less dramatic than could be anticipated.

In short, initial evidence seems to indicate that measurement equivalence between Web-based and paper-and-pencil tests is generally established. In addition, no large differences are found between supervised and unsupervised testing. Again, these results should be interpreted with caution because of the small number of research studies involved.

Applicant perceptions

Because test administration in an Internet-based environment differs from traditional testing, research has also begun to examine applicant reactions to Internet-based assessment systems. Mead (2001) reported that 81% of existing users were satisfied or quite satisfied with an on-line version of the 16PF Questionnaire. The most frequently cited advantage was the remote administration, followed by the quick reporting of results. The reported rate of technical difficulties was the only variable that separated satisfied from dissatisfied users. Another study by Reynolds et al. (2000) confirmed these results. They found more positive perceptions of actual applicants toward Internet-based testing than toward traditional testing. However, a confound was that all people receiving the Web-based testing format had opted for this format. Similar to Mead (2001), Reynolds et al. noted a heightened attention of applicants to technological and time-related factors (e.g., speed) when testing via the Internet as compared with traditional testing. No differences in applicant reactions across members of minority and non-minority groups were found.

Sinar and Reynolds (2001) conducted a multi-stage investigation of applicant reactions to supervised Internet-based selection procedures. Their

sample consisted of applicants for real job opportunities. They first gathered open-ended comments from applicants to Internet-based testing systems. About 70% of the comments obtained were positive. Similar to Reynolds et al. (2000), the speed and the efficiency of the Internet testing tool was the most important consideration of applicants, especially if the speed was slow. Many applicants also commented on the novelty of Internet-based testing. User-friendliness (e.g., ease of navigation) was another theme receiving substantial attention. Sinar and Reynolds also discovered that comments about user-friendliness, personal contact provided, and speed/efficiency were linked to higher overall satisfaction with the process. Finally, Sinar and Reynolds explored whether different demographic groups had different reactions to these issues. Markedly, there were more positive reactions for racial minorities, but user-friendliness discrepancies for females and older applicants. It is clear that more research is needed here to confirm and explain these findings.

In light of the aforementioned dimensions of Internet-based testing, a noteworthy finding of Sinar and Reynolds (2001) was that, on average, actual applicants reported a preference for the proctored (supervised) Web-based setting instead of taking the Web-based assessment from a location of their choice (unsupervised). Perhaps applicants considered the administrator's role to be crucial in informing applicants and providing help when needed. It is also possible that candidates perceived higher test security problems in the unsupervised Web-based environment.

Other research focused on the effects of different formats of Internet-based testing on perceptions of anonymity. However, a drawback is that this issue has only been investigated with student samples. Joinson (1999) compared socially desirable responding among students, who either completed personality-related questionnaires via the Web (unsupervised) or during courses (supervised). Both student groups were required to identify themselves (non-anonymity situation), which makes this experiment somewhat generalizable to a personnel selection context. Joinson found that responses of the unsupervised Web group exhibited significantly lower social desirability than people completing the questionnaires during supervised courses. He related this to the lack of observer presence inherent in unsupervised Internet-based testing. In a similar vein, Oswald et al. (2001) reported greater feelings of anonymity for completing personality measures in the Web/unsupervised condition vs. in the Web/supervised condition. Oswald et al. suggested that students probably felt more anonymous in the unsupervised setting because this setting was similar to surfing the Internet in the privacy of one's home.

Taken together, applicant perceptions of Internet-based testing applications seem to be favorable. Yet, studies also illustrate that demographic variables, technological breakdowns, and an unproctored test environment impact negatively on these perceptions.

Recommendations for Future Research

Given the state-of-the art of research on Internet testing applications, this last section proposes several recommendations for future research. In particular, we posit that future research should (1) learn from the lessons of the computer-based testing literature, (2) draw on psychological theories for examining Internet-enabled testing applications, and (3) address questions of most interest to practitioners.

Be aware of the lessons from computer-based testing research

As already mentioned, some Internet testing applications have a lot of similarities to computerized testing. Therefore, it is important that future research builds on this body of literature (Bartram, 1994; Burke, 1993; McBride, 1998, for reviews). Several themes may provide inspiration to researchers.

Measurement equivalence is one of the themes that received considerable attention in the computerized testing literature. On the one hand, there is evidence in the computerized testing literature that the equivalence of computerized cognitive ability measures to traditional paper-and-pencil measures is high. Mead and Drasgow's (1993) meta-analysis of cognitive ability measures found average cross-mode correlations of 0.97 for power tests. On the other hand, there is considerable debate whether computerized noncognitive measures are equivalent to their paper-and-pencil versions (King & Miles, 1995; Richman, Kiesler, Weisband, & Drasgow, 1999). This debate about the equivalence of noncognitive measures centers on the issue of social desirability. A first interpretation is that people display more candor and less social desirability in their responses to a computerized instrument. This is because people perceive computers to be more anonymous and private. Hence, according to this interpretation, they are more willing to share personal information. A second interpretation posits that people are more worried when interacting with a computer because they fear that their responses are permanently stored and can be verified by other parties at all times. In turn, this leads to less self-disclosure and more socially desirable responding. Recently, Richman et al. (1999) meta-analyzed previous studies on the equivalence of noncognitive measures. They also tested under which conditions computerized noncognitive measures were equivalent to their paper-and pencil counterparts. They found that computerization had no overall effect on measures of social desirability. However, they reported that being alone and having the opportunity to backtrack and to skip items resulted in more self-disclosure and less socially desirable responding among respondents. In more general terms, Richman et al. (1999) concluded that computerized questionnaires produced less social desirability when

participants were anonymous and when the questionnaire format mimicked that of a paper-and-pencil version.

Although researchers have begun examining the measurement equivalence issue in the context of Web-based testing, we believe that researchers may go even further because the computerized testing literature on measurement equivalence has important implications for future studies on Web-based testing. First of all, it does not suffice to examine measurement equivalence *per se*. The literature on computerized testing teaches us that it is crucial to examine under which conditions measurement equivalence is reduced or increased. Along these lines, several of the conditions identified by Richman et al. (1999) have direct implications for Web-enabled testing. For example, being alone relates to the Web-based test administration dimension 'no presence of test-administrator' that we discussed earlier. So, future research should examine the equivalence of Web-based testing under different test administration conditions. Especially lab research may be useful here. Second, the fact that Richman et al. (1999) found different equivalence results for noncognitive measures in the anonymous vs. the non-anonymous condition, calls for research in situations in which test results have consequences for the persons involved. Examples include field research with real applicants in actual selection situations or laboratory research in which participants receive an incentive to distort responses. Third, prior studies mainly examined the construct equivalence of Web-based tests. To date, no evidence is available as to how Internet-based administration affects the criterion-related validity of cognitive and noncognitive tests. Again, the answer here may depend on the type of Internet-based application.

Another theme from the computerized testing literature pertains to the impact of demographic variables on performance of computerized instruments (see Igbaria & Parasuraman, 1989, for a review). In fact, there is meta-analytic evidence that female college students have substantially more computer anxiety (Chua, Chen, & Wong, 1999) and less computer self-efficacy (Whitley, 1997) than males. Regarding age, computer confidence and control seems to be lower among persons above 55 years (Czaja & Sharit, 1998; Dyck & Smither, 1994). In terms of race, Badagliacco (1990) reported that whites had more years of computer experience than members of other races and Rosen, Sears, and Weil (1987) found that white students had significantly more positive attitudes toward computers. Research has begun to investigate the impact of these demographic variables in an Internet context. For example, Schumacher and Morahan-Martin (2001) found that males had higher levels of experience and skill using the Internet. In other words, as could be expected, the initial findings suggest that the trends found in the computerized testing literature (especially those with regard to gender and age) generalize to Internet-based applications. Although definitely more research is needed here, it is possible that Internet-enabled testing would

suffer from the so-called digital divide because some groups (females and older people) are disadvantaged in Internet testing applications. For practitioners, the future challenge then consists of implementing Internet tests that produce administrative and cost efficiencies and at the same time ensure fairness (Stanton & Rogelberg, 2001a). Researchers should study differential item/scale/test functioning across Web-based testing and traditional paper-and-pencil administrations. In addition, they should investigate which forms of Web-based testing produce less adverse impact. For example, it is likely that there is an interaction between the Web-based testing conditions and the occurrence of adverse impact. In particular, when organizations do not restrict the time and location of Web-based testing so that people can complete tests in each Internet-enabled terminal (e.g., in libraries, shopping centers), we expect that adverse impact against minority groups will be less as compared with proctored Web-based testing applications.

Finally, we believe that research on Web-based testing can learn from the history of computer-based testing. As reviewed by McBride (1998), the first wave of computerized testing primarily examined whether using a computerized administration mode was cost-efficient, whereas the second wave focused on converting existing paper-and-pencil instruments to a computerized format and studying measurement equivalence. According to McBride (1998), only the third wave of studies investigated whether a computerized instrument can actually change and enhance existing tests (e.g., by adding video, audio). Our review of current research shows that history seems to repeat itself. Current studies have mainly concentrated on cost savings and measurement equivalence. So, future studies are needed that examine how use of the Internet can actually change the actual test and the test administration process.

Draw on psychological theories to examine Internet-based testing applications

Our review of current research on Internet-based testing illustrated that few studies were grounded on a solid theoretical framework. However, we believe that at least the following two theories may be fruitfully used in research on Internet-based testing, namely organizational privacy theory and organizational justice theory. Although both theories are related (see Bies, 1993; Eddy, Stone, & Stone, 1999; Gilliland, 1993), we discuss their potential benefits in future research on Internet-based testing separately.

Organizational privacy theory (Stone & Stone, 1990) might serve as a first theoretical framework to underpin research on Internet-based testing applications. Privacy is a relevant construct in Internet-based testing because of several reasons. First, Internet-based testing applications are typically non-anonymous. Second, applicants are often asked to provide personal and sensitive information. Third, applicants know that the information is

captured in electronic format, facilitating multiple transmissions over the Internet and storage in various databases (Stanton & Rogelberg, 2001a). Fourth, privacy concerns might be heightened when applicants receive security messages (e.g., secure server probes, probes for accepting cookies, etc.). Although Bartram (2001) argued that these security problems are largely overstated, especially people who lack Internet experience, Internet self-efficacy, or the belief that the Internet is secure may worry about them.

So far, there has been no empirical research on the effects of Web-based testing on perceptions of invasion of privacy. Granted, there is some evidence that people are indeed more wary about privacy when technology comes into play, but these were purely descriptive studies. Specifically, Eddy et al. (1999) cited several surveys that found that public concern over invasion of privacy was on the rise. They linked this increased concern over privacy to the recent technological advances that have occurred. Additionally, Cho and LaRose (1999) cited a survey in which seven out of ten respondents to an on-line survey worried more about privacy on the WWW than through the mail or over the telephone (see also Hoffman, Novak, & Peralta, 1999; O'Neil, 2001).

In the privacy literature, there is general consensus that privacy is a multi-faceted construct. For example, Cho and LaRose (1999) made a useful distinction between physical privacy (i.e., solitude), informational privacy (i.e., the control over the conditions under which personal data are released), and psychological privacy (i.e., the control over the release of personal data). In a similar vein, Stone and Stone (1990) delineated three main themes in the definition of privacy. A first form of privacy is related to the notion of information control, which refers to the ability of individuals to control information about them. This meaning of privacy is related to the psychological privacy construct of Cho & LaRose (1999). Second, Stone and Stone (1990) discuss privacy as the regulation of interactions with others. This form of privacy refers to personal space and territoriality (cf. the physical privacy of Cho and LaRose, 1999). A third perspective on privacy views it in terms of freedom from the influence or control by others (Stone & Stone, 1990).

Several studies in the privacy literature documented that especially the perceived control over the use of disclosed information is of pivotal importance to the notion of invasion of privacy (Fusilier & Hoyer, 1980; Stone, Gueutal, Gardner, & McClure, 1983; Stone & Stone, 1990). This perceived control is typically broken down into two components, namely the ability to authorize disclosure of information and the target of disclosure. There is also growing support for these antecedents in the context of the use of information technology. For instance, Eddy et al. (1999) examined reactions to human resource information systems and found that individuals perceived a policy to be most invasive when they had no control over the release of personal information *and* when the information was provided to parties outside the organization.

How can privacy theory advance our understanding of applicants' view of Web-based testing? First, we need a clear understanding of which forms of privacy are affected by Web-based testing. Studies are needed to examine how the different forms of Web-based testing outlined above affect the various forms of privacy. Second, we need studies to shed light into the antecedents of applicants' privacy concerns. Studies are needed to confirm whether applicants' perceived decrease of control over the conditions under which personal information might be released and over the organizations that subsequently might use it are the main determinants to trigger privacy concerns in Web-based testing applications. Again, it would be interesting to examine this for the various dimensions of Web-enabled testing. Third, future studies should examine the consequences of applicants' privacy concerns in web-based testing. For example, when privacy concerns are heightened, do applicants engage in more socially desirable responding and less self-disclosure? What are the influences on their perceptions of the Web-based testing application? A final avenue for future research consists of investigating under which conditions these privacy concerns might be reduced or alleviated. To this end, research could manipulate the various dimensions of Web-based testing and examine their impact on different forms of privacy. For example, does a less restrictive interface reduce privacy concerns? Similarly, how do technology and disclaimers that guarantee security and confidentiality affect applicants' privacy concerns? What roles do the type of test and the kind of information provided play? What is the influence of the presence of a test-administrator? As mentioned above, there is some preliminary evidence that applicants feel more (physical) privacy and provide more candid answers when no test-administrator is present (see Joinson, 1999; Oswald et al., 2001). Such studies might contribute to our general understanding of privacy in technological environments but might also provide concrete recommendations for improving current Web-based testing practices.

A second theoretical framework that may be relevant is organizational justice theory (Gilliland, 1993; Greenberg, 1990). Organizational justice theory in general and a justice framework applied to selection in particular are relevant here because applicants are likely to compare the new Web-based medium with more traditional approaches. Hence, one of applicants' prime concerns will be whether this new mode of administration is more or less fair than the traditional ones. Gilliland (1993) presented a model that integrated both organizational justice theory and prior applicant reactions research. Two central constructs of the model were distributive justice and procedural justice, which both had their own set of distinct rules (e.g., job-relatedness, consistency, feedback, two-way communication). Gilliland (1993) also delineated the antecedents and consequences of possible violation of these rules.

Here we only discuss the variables that may warrant special research

attention in the context of Web-based testing. First, Gilliland's (1993) model should be broadened to include technological factors as possible determinants of applicants' fairness reactions. As mentioned above, initial research on applicant reactions to Web-based testing suggests that these reactions are particularly influenced by technological factors such as slowdowns in the Internet connection or Internet connection crashes. Apparently, applicants expect these technological factors to be flawless. If the technology fails for some applicants and runs perfectly for others, fairness perceptions of Web-based testing are seriously affected. Gilliland's (1993) model should also be broadened to include specific determinants to computerized/Web-based forms of testing such as Internet/computer anxiety and Internet/computer self-efficacy. Second, Web-based testing provides excellent opportunities for testing an important antecedent of the procedural justice rules outlined in Gilliland's (1993) model, namely the role of 'human resource personnel' (e.g., test-administrators). As mentioned above, in some applications of Web-based testing, the role of test-administrators is reduced or even discarded. The question remains how this lack of early stage face-to-face contact (one of Gilliland's, 1993, procedural justice rules) affects applicants' reactions during and after hiring (Stanton & Rogelberg, 2001a). On the one hand, applicants might perceive the Web-based testing situation as more fair because the user interface of a computer is more neutral than a test administrator. On the other hand, applicants might regret that there is no 'live' two-way communication, although many user interfaces are increasingly interactive and personalized. An examination of the effects of mixed mode administration might also clarify the role of test-administrators in determining procedural justice reactions. Mixed mode administration occurs when some tests are administered via traditional means, whereas other tests are administered via the Internet. We believe that these different modalities of Web-based testing offer great possibilities for studying specific components of Gilliland's justice model and for contributing to the broader justice literature. Third, in current Web-based testing research, applicants' reactions to Web-based testing are hampered by the 'novelty' aspect of the new technology. This novelty aspect creates a halo effect so that it is difficult to get a clear insight into the other bases of applicants' reactions to Web-based testing applications. Therefore, future research should pay particular attention to one of the moderators of Gilliland's model, namely applicants' prior experience. In the context of Web-based testing, this moderator might be operationalized as previous work experience in technological jobs or prior experience with Internet-based recruitment/testing.

Address questions of interest to practitioners

The growth of Internet-based testing opens a window of opportunities for researchers as many organizations are asking for suggestions and advice.

Besides answering questions that are consistent with previous paradigms (see our first two recommendations), it is equally important to examine the questions that are on the top of practitioners' minds.

When we browsed through the popular literature on Internet-based testing, cost benefits and concerns definitely emerged as a prime issue. To date, most practitioners are convinced of the possible benefits of Internet recruitment. Similarly, there seems to exist general consensus that Internet testing may have important advantages over paper-and-pencil testing. This is also evidenced by case studies. Baron and Austin (2000) reported results of a case study in which an organization (America Online) used the Internet for screening out applicants in early selection stages. They compared the testing process before and after the introduction of the Internet-based system on a number of ratios. Due to Internet-enabled screening the time per hire decreased from 4 hours and 35 minutes to 1 hour and 46 minutes so that the whole process was reduced by 20 days. Sinar and Reynolds (2001) also referred to case studies that demonstrated that companies can achieve hiring cycle time reductions of 60% through intensive emphasis on Internet staffing models. A limitation of these studies, however, is that they evaluated the combined impact of Internet-enabled recruitment and testing (i.e., screening), making it difficult to understand the unique impact of Internet testing.

More skepticism, though, surrounds the incremental value of supervised Internet-based testing over 'traditional' computerized testing within the organization. In other words, what is the added value of having applicants complete the tests at various test centers vs. having them complete tests in the organization? The obvious answer is that there is increased flexibility for both the employer and the applicant and that travel costs are reduced. Yet, not everybody seems to be convinced of this. A similar debate exists about the feasibility of having an unproctored Web-based test environment (in terms of user identification and test security). Therefore, future studies should determine the utility of various Web-based testing applications and formats. To this end, various indices can be used such as time and cost savings and applicant reactions (Jayne & Rauschenberger, 2000).

A second issue emerging from popular articles about Internet selection processes relates to practitioners' interest as to whether use of Web-based testing has positive effects on organizations' general image and their image as employers particularly. Although no studies have been conducted, prior studies in the broader selection domain support the idea that applicants' perceptions of organizational image are related to the selection instruments used by organizations (e.g., Macan, Avedon, Paese, & Smith, 1994; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). Moreover, Richman-Hirsch, Olson-Buchanan, and Drasgow (2000) found that an organization's use of multimedia assessment for selection purposes might signal something about an organization's technological knowledge and savvy. Studies are needed to

confirm these findings in the context of Web-based testing. Again, attention should be paid here to the various dimensions (especially technology and possible technological failures) of Web-based testing as potential moderators of the effects.

CONCLUSION

The aim of this chapter was to review existing research on Internet recruitment and testing and to formulate recommendations for future research. A first general conclusion is that research on Internet recruitment and testing is still in its early stages. This is logical because of the relatively recent emergence of the phenomenon. Because only a limited number of topics have been addressed, many issues are still open. As noted above, the available studies on Internet recruitment, for example, have mainly focused on applicant (student) reactions to the Internet as a recruitment source, with most studies yielding positive results for the Internet. However, key issues such as the decision-making processes of applicants and the effects of Internet recruitment on post-recruitment variables such as company image, satisfaction with the selection process, or withdrawal from the current organization have been ignored so far.

As compared with Internet-based recruitment, more research attention has been devoted to Internet testing. Particularly, measurement equivalence and applicant reactions have been studied, with most studies yielding satisfactory results for Internet testing. Unfortunately, some crucial issues remain either unresolved (i.e., the effects of Internet testing on adverse impact) or unexplored (i.e., the effects of Internet testing on criterion-related validity).

A second general conclusion is the lack of theory in existing research on Internet testing and recruitment. To this end, we formulated several suggestions. As noted above, we believe that the elaboration likelihood model and resource exchange theory may be fruitfully used to understand Internet job site choice better. We have also advocated that organizational privacy theory and organizational justice theory might advance existing research on Internet testing.

Finally, we acknowledge that it is never easy to write a review of an emerging field such as Internet-based recruitment and testing because at the time this chapter goes to print, new developments and practices will have found inroad in organizations and new research studies will have been conducted. Again, this shows that for practitioners and researchers the application of new technologies such as the Internet to recruitment and testing is both exciting and challenging.

AKNOWLEDGEMENTS

This research was supported by a grant as Postdoctoral Fellow of the Fund for Scientific Research—Flanders (Belgium) (F.W.O.—Vlaanderen). We would like to thank Fred Oswald and Frederik Anseel for their comments on an earlier version of this chapter.

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Badagliacco, J. M. (1990). Gender and race differences in computing attitudes and experience. *Social Science Computer Review*, **8**, 42–63.

Barber, A. E. (1998). *Recruiting Employees: Individual and Organizational Perspectives*. Thousands Oaks, CA: Sage.

Barber, A. E., Daly, C. L., Giannantonio, C. M., & Phillips, J. M. (1994). Job search activities: An examination of changes over time. *Personnel Psychology*, **47**, 739–766.

Baron, H., & Austin, J. (2000). Measuring ability via the Internet: Opportunities and issues. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Bartram, D. (1994). Computer-based assessment. In C. L. Cooper, & I. T. Robertson (eds), *International Review of Industrial and Organizational Psychology* (Vol. 9, pp. 31–69). New York: John Wiley & Sons.

Bartram, D. (2001). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, **8**, 261–274.

Beaty, J. C., Jr, Fallon, J. D., & Shepherd, W. (2002, April). Proctored versus unproctored web-based administration of a cognitive ability test. In F. L. Oswald, & J. M. Stanton (chairs), Being virtually hired: Implications of web testing for personnel selection. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Bies, R. J. (1993). Privacy and procedural justice in organizations. *Social Justice Research*, **6**, 69–86.

Boehle, S. (2000). Online recruiting gets sneaky. *Training*, **37** (May), 66–74.

Breaugh, J., & Starke, M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of Management*, **26**, 405–434.

Brinberg, D., & Ganesan, S. (1993). An application of Foa's resource exchange theory to product positioning. In U. Foa, J. Converse, K. Y. Tornblom, & E. Foa (eds), *Resource Theory: Explorations and Applications* (pp. 219–231). San Diego, CA: Academic Press.

Brooks, P. W., Jr (2000, June). Internet assessment: Opportunities and challenges. Paper presented at the 24th Annual IPMAAC Conference on Professional Personnel Assessment, Washington, DC.

Buchanan, T., & Smith, J. L. (1999a). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, **90**, 125–144.

Buchanan, T., & Smith, J. L. (1999b). Research on the internet: Validation of a World-Wide Web mediated personality scale. *Behavior Research Methods, Instruments, & Computers*, **31**, 565–571.

Burke, M. J. (1993). Computerized psychological testing: Impacts on measuring predictor constructs and future job behavior. In N. Schmitt, & W. C. Borman (eds), *Personnel Selection in Organizations* (pp. 203–239). San Francisco: Jossey-Bass.

Burnkrant, S., & Taylor, C. D (2001, April). Equivalence of traditional and Internet-based data collection: Three multigroup analyses. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Cappelli, P. (2001, March). Making the most of on-line recruiting. *Harvard Business Review*, **79**, 139–146.

Cho, H., & LaRose, R. (1999). Privacy issues in Internet surveys. *Social Science Computer Review*, **17**, 421–434.

Chua, S. L., Chen, D., & Wong, A. F. L. (1999). Computer anxiety and its correlates: A meta-analysis. *Computers in Human Behavior*, **15**, 609–623.

Cober, R. T., Brown, D. J., Blumental, A. J., Doverspike, D., & Levy, P. (2000). The quest for the qualified job surfer: It's time the public sector catches the wave. *Public Personnel Management*, **29**, 479–494.

Cober, R. T., Brown, D. J., Blumental, A. J., & Levy, P. E. (2001, April). What do the 'Best Companies' do? A qualitative analysis of internet recruiting practices. In S. Highhouse (chair), New approaches to research on job search and job choice. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Coffee, K., Pearce, J., & Nishimura, R. (1999). State of California: Civil service testing moves into cyberspace. *Public Personnel Management*, **28**, 283–300.

Czaja, S., & Sharit, J. (1998). Age differences in attitudes toward computers. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences*, **53**, 329–340.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments, & Computers*, **31**, 572–577.

Dineen, B. R., Ash, S. R., & Noe, R. A. (2002). A web of applicant attraction: Person–organization fit in the context of web-based recruitment. *Journal of Applied Psychology* **87**, 723–734.

Dyck, J. L., & Smither, J. A. (1994). Age differences in computer anxiety: The role of computer experience, gender and education. *Journal of Educational Computing Research*, **10**, 239–248.

Eddy, E. R., Stone, D. L., & Stone, E. F. (1999). The effects of information management policies on reactions to human resource information systems: An integration of privacy and procedural justice perspectives. *Personnel Psychology*, **52**, 335–358.

Elgin, P. D., & Clapham, M. M. (2000). Attributes associated with electronic versus paper resumes. Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Fenlason, K. J. (2000, April). Multiple data collection methods in 360-feedback programs: Implication for use and interpretation. Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Fisher, C. D., Ilgen, D. R., Hoyer, W. D. (1979). Source credibility, information favorability, and job offer acceptance. *Academy of Management Journal*, **22**, 94–103.

Foa, U., Converse, J., Tornblom, K. Y., & Foa, E. (1993). *Resource Theory: Explorations and Applications*. San Diego, CA: Academic Press.

Fusilier, M. R., & Hoyer, W. D. (1980). Variables affecting perceptions of invasion of privacy in a personnel selection situation. *Journal of Applied Psychology*, **65**, 623–626.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, **18**, 694–734.

Greenberg, C. I. (1999). Technological innovations and advancements for psychologists working with organizations. *The Psychologist-Manager Journal*, **3**, 181–190.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management*, **16**, 399–342.

Gutmacher, G. (2000). Secrets of online recruiters exposed! *Workforce*, **79** (October), 44–50.

Harris, M. M. (1999, April). Practice network: I-O psychology.com-the internet and I-O psychology. *The Industrial-Organizational Psychologist*, **36**, 89–93.

Harris, M. M. (2000). The Internet and industrial/organizational psychology: Practice and research perspectives. *Journal of e-Commerce and Psychology*, **1**, 4-23.

Harris, M. M., & DeWar, K. (2001, April). Understanding and using Web-based recruiting and screening tools: Key criteria, current trends, and future directions. Workshop presented at the Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Hezlett, S. A. (2000, April). Employee attitude surveys in multi-national organizations: An investigation of measurement equivalence. Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Highhouse, S., & Hoffman, J. R. (2001). Organizational attraction and job choice. In C. L. Cooper, & I. T. Robertson (eds), *International Review of Industrial and Organizational Psychology* (pp. 37–64). Chichester, UK: John Wiley & Sons.

Highhouse, S., Stanton, J. M., & Reeve, C. L. (Forthcoming). Asymmetries in reactions to positive and negative information in an organizational recruitment context. *Journal of Vocational Behavior*.

Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM*, **42**, 80–85.

Hom, P. W., & Griffeth, R. W. (1995). *Employee Turnover*. Cincinnati, OH: South-Western.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—remembering the past. *Annual Review of Psychology*, **51**, 631–664.

Igbaria, M., & Parasuraman, S. (1989). A path analytic study of individual characteristics, computer anxiety and attitudes towards microcomputers. *Journal of Management*, **15**, 373–388.

Jayne, M. E. A., & Rauschenberger, J. M. (2000). Demonstrating the value of selection in organizations. In J. F. Kehoe (ed.), *Managing Selection in Changing Organizations* (pp. 123–157). San Francisco: Jossey-Bass.

Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, **31**, 433–438.

King, W. C., & Miles, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology*, **80**, 643–651.

Larsen, D. A., & Phillips, J. I. (2001). Effect of recruiter on attraction to the firm: Implications of the elaboration likelihood model. *Journal of Business and Psychology*, **16**, 347–364.

Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *Academy of Management Review*, **19**, 51–89.

Lent, R. W. (2001). Vocational psychology and career counseling: Inventing the future. *Journal of Vocational Behavior*, **59**, 213–225

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, **47**, 715–738.

Magnan, S. M., Lundby, K. M., & Fenlason, K. J. (2000, April). Dual media: The art and science of paper and internet employee survey implementation. Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Maher, K., & Silverman, R. E. (2002, January). Online job sites yield few jobs, users complain. *The Wall Street Journal*, January 2, A1, A13.

McBride, J. R. (1998). Innovations in computer-based ability testing: Promise, problems and perils. In M. D. Hakel (ed.), *Beyond Multiple Choice* (pp. 113–129). Mahwah, NJ: Lawrence Erlbaum.

Mead, A. D. (2001, April). How well does web-based testing work? Results of a survey of users of NetAssess. In F. L. Oswald (chair), Computers = good? How test-user and test-taker perceptions affect technology-based employment testing. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Mead, A. D., & Coussons-Read, M. (2002, April). The equivalence of paper- and web-based versions of the 16PF questionnaire. In F. L. Oswald, & J. M. Stanton (chairs), Being virtually hired: Implications of web testing for personnel selection. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, **114**, 449–458.

Mitchell, T. R., Holtom, B. C., & Lee, T. W. (2002). How to keep our best employees: Developing an effective retention policy. *Academy of Management Executive*, **15**, 96–108.

Newhagen, J. E., & Rafaeli, S. (2000). Why communication researches should study the Internet: A dialogue. *Journal of Communication*, **46**, 4–13.

Oliver, L. W., & Whiston, S. C. (2000). Internet career assessment for the new millennium. *Journal of Career Assessment*, **8**, 361–369.

O'Neil, D. (2001). Analysis of Internet users' level of online privacy concerns. *Social Science Computer Review*, **19**, 17–31.

Oswald, F. L., Carr, J. Z., & Schmidt, A. M. (2001, April). The medium and the message: Dual effects of supervision and web-based testing on measurement equivalence for ability and personality measures. In F. L. Oswald (chair), Computers = good? How test-user and test-taker perceptions affect technology-based employment testing. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Pasveer, K. A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavior Research Methods, Instruments, & Computers*, **30**, 309–313.

Pettit, F. A. (1999). Exploring the use of the World Wide Web as a psychology data collection tool. *Computers in Human Behavior*, **15**, 67–71.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.

Ployhart, R. E., Weekley, J., Holtz, B., & Kemp, C. (2002, April). Web-based vs. paper and pencil testing: A comparison of factor structures across applicants and incumbents. In F. L. Oswald, & J. M. Stanton (chairs), Being virtually hired: Implications of web testing for personnel selection. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Reynolds, D. H., Sinar, E. F., & McClough, A. C. (2000, April). Evaluation of an Internet-based selection procedure. In N. J. Mondragon (chair), Beyond the demo: The empirical nature of technology-based assessments. Symposium presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Richman, W.L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, **84**, 754–775.

Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, **85**, 880–887.

Rosen, L. D., Sears, D. C., & Weil, M. M. (1987). Computerphobia. *Behavior Research Methods, Instruments, & Computers*, **19**, 167–179.

Rozelle, A. L., & Landis, R. S. (2002). An examination of the relationship between use of the Internet as a recruitment source and student attitudes. *Computers in Human Behavior*, **18**, 593–609.

Saks, A. M., & Ashforth, B. E. (2000). Change in job search behaviors and employment outcomes. *Journal of Vocational Behavior*, **56**, 277–287.

Salgado, J. F. (1999). Personnel selection methods. In C. L. Cooper, & I. T. Robertson (eds), *International Review of Industrial and Organizational Psychology* (Vol. 14, pp. 1–54). Chichester, UK: John Wiley & Sons.

Scheu, C., Ryan, A. M., & Nona, F. (1999, April). Company web sites as a recruiting mechanism: What influences applicant impressions? Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Schmitt, N., & Chan, D. (1998). *Personnel Selection: A Theoretical Approach*. Thousand Oaks, CA: Sage.

Schumacher, P., & Morahan-Martin, J. (2001). Gender, Internet and computer attitudes and experiences. *Computers in Human Behavior*, **17**, 95–110.

Sharf, J. (2000). As if 'g-loaded' adverse impact isn't bad enough, Internet recruiters can expect to be accused of 'e-loaded' impact. *The Industrial-Organizational Psychologist*, **38**, 156.

Silverman, R. E. (2000). Raiding talent via the Web. *The Wall Street Journal*, October 3, B1, B16.

Simsek, Z., & Veiga, J. F. (2001). A primer on Internet organizational surveys. *Organizational Research Methods*, **4**, 218–235.

Sinar, E. F., & Reynolds, D. H. (2001, April). Applicant reactions to Internet-based selection techniques. In F. L. Oswald (chair), Computers = good? How test-user and test-taker perceptions affect technology-based employment testing. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Smith, M. R., Rogg, K. L., & Collins P. (2001, April). Applying Internet-based business simulations to individual assessment: Some results and implications. In M. L. Kelly, & M. T. Russell (chairs), *The use of assessment tools in leadership development*. Practitioner forum presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, **46**, 49–76.

Spera, S. D., & Moye, N. A. (2001, April). Measurement equivalence between paper and web survey methods in a multinational company. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology*, **51**, 709–725.

Stanton, J. M. (1999). Validity and related issues in Web-based hiring. *The Industrial-Organizational Psychologist*, **36**(3), 69–77.

Stanton, J. M., & S. G. Rogelberg, S.G. (2001a, April). Challenges and obstacles in conducting employment testing via the Internet. In F. L. Oswald, (chair), Computers = good? How test-user and test-taker perceptions affect technology-based employment testing. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Stanton, J. M., & Rogelberg, S. G. (2001b). Using Internet/Intranet web pages to collect organizational research data. *Organizational Research Methods*, **4**, 199–216.

Stone, E. F., Gueutal, H. G., Gardner, D. G., & McClure, S. (1983). A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations. *Journal of Applied Psychology*, **68**, 459–468.

Stone, D. L., & Stone, E. F. (1990). Privacy in organizations: Theoretical issues, research findings, and protection mechanisms. *Research in Personnel and Human Resources Management*, **8**, 349–411.

Straus, S. G., Miles, J. A., & Levesque, L. L. (2001). The effects of videoconference, telephone, and face-to-face media on interviewer and applicant judgments in employment interviews. *Journal of Management*, **27**, 363–381.

Travagline, A., & Frei, R. L. (2001, April). 21st century recruiting: A model for Internet-based realistic job previews. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Weiss, E. M., & Barbeite, G. F. (2001). Internet as a job source and job site preference. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Whitley, B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior*, **13**, 1–22.

Zottoli, M. A., & Wanous, J. P. (2000). Recruitment source research: Current status and future directions. *Human Resource Management Review*, **10**, 353–382.

Zusman, R. R., & Landis, R. (2002). Applicant preferences for Web-based versus traditional job postings. *Computers in Human Behavior*, **18**, 285–296.

Chapter 5

WORKAHOLISM: A REVIEW OF THEORY, RESEARCH, AND FUTURE DIRECTIONS

Lynley H. W. McMillan and Michael P. O'Driscoll
University of Waikato

and Ronald J. Burke
York University

Workaholism involves difficulty disengaging from work, a strong drive to work, intense enjoyment of work, and a differing use of leisure time than others. As both work and leisure trace to the heart of our kinship with other humans, their struggle for primacy is a historical one that has its roots as early as the 14th century. Until this time, with the exception of the Roman era, people generally worked until they had enough food and then rested and played in the remainder of their time (Preece, 1981). Thus, the nature and structure of the working day varied, depending on the season, the weather, and the availability of food. Families worked as units that comprised several generations, from the very small and very elderly (who contributed as best they could) to the physically able (who carried most of the workload). In 1335, however, the invention of the mechanical clock provided an independent measure of people's working hours. This catalysed a move away from the cycles of nature (where work and leisure were intertwined) to an artificial dichotomy of 'work' and 'leisure'.

The following half-century brought the advent of cloth manufacture, the birth of industry, the invention of 'fashion', and the start of contract labour, with many families 'putting out' (contracting their services to the textile industry: Preece, 1981). By the time the printing press was developed in 1450, employers had begun to push for 12-hour working days, with the less scrupulous among them hiding clocks to surreptitiously extract more

---

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

working time. The Protestants, who disparaged luxury and exalted hard work, supported this new 'mean-ness' with time. By 1600, the arrival of the British middle class sparked a new demand for leisure and extravagance. Soon thereafter the first factories opened and the Puritans began a crusade to 'purge the disorder of leisure' from the world. And so the battle between work and leisure oscillated for another 150 years (Cross, 1990).

Paradoxically, as the Industrial Revolution began in 1780, the concept of holiday resorts took hold, forging a wider chasm between work and leisure. In 1866 employees fought for an 8-hr day, but the introduction of the light bulb in 1880 perpetuated employers' override of nature's patterns and enabled a 24-hr working day. Thus, it was not until 1938 that the general workforce was allowed weekends, paid holidays, and a 40-hour week (Robinson & Godbey, 1992). Subsequently, a new 'time consciousness' evolved and catalysed an avalanche of time-saving technological inventions, a rush toward the cities, and a new flush of spending. In the mid-1960s, however, a critical juncture occurred; while many parents continued to march to the beat of consumerism, many of their children began to join the antithetical hippie movement that rejected the parental work ethic in favour of 'lifestyle.' It is against this backdrop of societal vacillation between valuing work and alternately leisure that in 1968 the word 'workaholism' evolved.

As technological inventions such as mobile phones, computers, faxes, and emails continued to mobilize the workforce, the boundary between work and home blurred and workaholism gained prominence in the public arena. Today, in addition to a plethora of media articles, there are multitudinous websites specific to workaholism, Workaholics Anonymous groups, residential treatment centres, books, therapists, and counselors that purport cures for workaholism. Thus, a ready audience of research consumers and stakeholders exists. Employers and organizational consultants are curious about the organizational value of workaholism, therapists are interested in how it is measured and treated, and the working public is keen to maximize benefits and minimize costs. Paradoxically, however, international communication and globalization of culture have only recently brought workaholism to the attention of academic researchers.

Originally, the word 'workaholism' was a take on working too hard in an alcoholic-like manner and was intended to connate all the problems that addiction brings (Oates, 1968). However, to this day, while most academics agree that work is healthy, desirable, and in fact protective from many illnesses, debate has continued over the merits and demerits of workaholism. Early research suggested that it was desirable (Machlowitz, 1978), but later studies disagreed (Robinson, 1996a), although most contemporary researchers agree that workaholism has two, possibly three components: (i) enjoyment, (ii) drive, and (iii) work involvement. However, some argue that this last factor saturates the other two, and is therefore redundant.

The importance of measurement validation has been one that has plagued the early development of workaholism research and substantially restricted generality. Currently, there are three validated measures of workaholism; the oldest is the Work Addiction Risk Test (WART: Robinson, 1998c), a family therapy-based, 25-item measure that assumes an addiction paradigm. The WART targets predominantly Type-A behaviour (i.e., life in general, as opposed to work-specific) and appears to be reliable, although validation studies have comprised restricted populations and thus validity is not convincingly established. The second measure is the *Schedule for Non Adaptive and Adaptive Personality Workaholism Scale* (SNAP-Work: Clark, McEwen, Collard, & Hickok, 1993), an 18-item instrument that assumes a degree of overlap with obsessive–compulsive personality disorder. The scale has high internal consistency, good split-half reliability, and demonstrates convergence with an alternate measure of workaholism (McMillan, O'Driscoll, Marsh, & Brady, 2001).

However, the most widely utilized instrument is the Workaholism Battery (WorkBAT: Spence and Robbins, 1992), a 25-item, self-report questionnaire comprising three scales; drive, work enjoyment, and work involvement. While the WorkBAT has relatively convincing content, face, and convergent validity (cf., Burke, 1999e), controversy exists over the internal factor structure. The first two scales have replicated in three separate factor analyses and have repeatedly demonstrated acceptable alpha coefficients across a broad range of populations (McMillan, Brady, O'Driscoll, & Marsh, 2002). Conversely however, the work involvement scale appears more problematic; three separate factor analyses have not replicated the factor (Kanai, Wakabayashi, & Fling, 1996; McMillan et al., in press). The measure has subsequently been revised to a 2-scale, 14-item instrument (the WorkBAT-R: McMillan et al., in press) that holds promising consistency, reliability, convergent validity, and scientific utility.

Taking the research field in a new direction, and on the presumption that the WorkBAT measures attitude and affect rather than overt behaviour, Mudrack and Naughton (2001) recently developed two new scales to measure the behavioural aspects of workaholism. They proposed that workaholism comprised two key elements: non-required work and interpersonal control. A confirmatory factor analysis supported the structure of the measure, while relations with the external criterion supported the empirical utility of the scales, although some qualifications apply. This sample worked excessive hours, were well educated, 46% had management-type roles and therefore more likely to assert control at work, and methodologically independent criteria were not used (e.g., direct observation). Thus, a promising start was made, but further validation is required.

While research interest in workaholism has mushroomed over the last five years, providing more incisive data and rigorous analyses, most of the literature remains dispersed between multiple disciplines and poorly integrated

into theoretical frameworks. While these weaknesses are inherent in all new research endeavours, the recent surge in academic interest and resultant publications suggest it is timely to adopt a more coherent, rigorous, and theoretically based approach to workaholism research. The present chapter therefore presents (i) a review and critique of theoretical models of workaholism, (ii) a summary of contemporary research, and (iii) a reflection on methodological and theoretical frameworks from which to conduct further research in this domain.

THEORIES OF WORKAHOLISM

To date, with the possible exception of some family-systems work, the majority of workaholism research has occurred from a wide variety of paradigms on an *ad hoc* basis without explication of a corresponding theory. The paradigms employed to date include addiction models, learning theory, trait-based paradigms, and, more recently, cognitive frameworks, along with family-systems models. Given the importance of clear theoretical frameworks in interpreting existing data and generating new hypotheses, the present section will expound these five models.

Addiction Theory

While the majority of workaholism research has implicated addiction as a causal factor (cf., Porter, 1996, for a precis of alcoholism–workaholism parallels), it has not directly linked the resultant data to *theory*. However, there are two broad classes of addiction theory that could be applied to workaholism data: the medical model and the psychological model (Eysenck, 1997).

Medical model of addiction

The medical model predicts that addiction occurs when a person becomes physically addicted to chemicals that are exogenous (e.g., drugs and alcohol), or endogenous (e.g., dopamine: Di Chiara, 1995). Some researchers have hypothesized that working long hours produces excessive adrenaline (Fassel, 1992). Adrenaline produces pleasurable somatic sensations and in turn becomes addictive, spurs the person to work more to produce more, and perpetuates an ongoing cycle of addiction (Fassel, 1992). Given the multifarious variables that produce adrenaline, however (e.g., racing for a deadline), statistically eliminating these mediating variables would be extremely complex and require highly technical biological tests. Unfortunately, while appropriate blood and urine tests are available, they are also open to confounding by the physiological stimulus of taking blood, dietary intake,

and circadian rhythms (Di Chiara, 1995). Despite this, authors continue to parallel workaholism with the 'classic' biological addiction symptoms of tolerance, craving, and withdrawal (cf., Robinson, 1998c). While the medical model provides an invitingly simple conceptualization of workaholism, and could be easily verified using a baseline and alternating treatment design with independent observers, none of the addiction hypotheses have been tested. Thus, the prerequisite for accepting the theory (i.e., a body of supporting data) has not been met.

Psychological model of addiction

The psychological model of addiction predicts that substance abuse continues despite having overt and sometimes distal disadvantages, as it confers some immediate benefits (Eysenck, 1997). Thus, people believe that they cannot function without these repetitive cycles of behaviour, and psychological dependence develops. This implies that workaholics perceive some degree of benefit (e.g., prestige) in perpetually working (Rohrlich, 1980) despite negative side-effects (e.g., tiredness). However, the model also implies that if prestige could be 'earned' by an alternate behaviour (such as coaching a sports team), then the alternate behaviour may become the focus of addiction instead. Thus, workaholism could be replaced by more adaptive behaviours.

Features, measures, and limitations of addiction theories. The current dearth of empirical data makes developing a comprehensive addiction theory of workaholism premature, especially as methodological complexities hinder progress and there are no relevant measures available. Medical theories are constrained by the fact that the addictive substance (work-generated adrenaline) is not as easy to isolate and measure as other addictive chemicals (drugs and alcohol). Additionally, workaholism does not appear to be surrounded by crime, street life, and 'user' cultures that are characteristic of other addictions (McMillan et al., 2001). Addiction theory, therefore, generates useful questions and hypotheses about the nature of workaholism, but, until more empirical data emerge, is unable to be verified.

Learning Theory

Of the three models inherent in learning theory (classical conditioning, social learning theory, and operant learning), operant learning is most relevant to workaholism. Operant learning predicts workaholism to be a relatively durable behaviour that is established through operant conditioning when a voluntary response comes under the control of its consequences by earning a desired outcome (Skinner, 1974). Thus, workaholism would arise after voluntarily working a few extra hours that led to pleasant peer approval and further increased the likelihood of discretionary working. While the

positive reinforcer (i.e., maintaining factor) in the present example is pleasant, this does not necessarily need to be the case. For instance, a negative reinforcer (escape from an unpleasant event such as conflict at home) may also maintain discretionary working. This conceptualization also has interesting links with compensation (e.g., non-work activities are on relatively lean reinforcement schedules) and spillover (e.g., busy-ness generalizes from home into the workplace). Alternatively, workaholism may arise from smaller behavioural repertoires (e.g., postponed gratification) that generalize into the workplace. Overall, operant conditioning implies that workaholism develops where it leads to desired outcomes, and thus dominates high-earning, high-status jobs, especially where home or leisure are unsatisfying. Most controversially, the theory predicts that workaholism could be shaped into anyone given adequately potent and idiopathically suitable reinforcers. It also implies that workaholism could be faded out of a person's behavioural repertoire (McMillan et al., 2001).

Features, measures, and limitations of learning theory. Learning theories are distinguished by their inherent optimism that workaholism could be readily trained out of people. Currently, there are no measures of workaholism that relate directly to learning theory. Because the theory avoids invoking reified explanatory fictions (e.g., personality) that are not directly observable, it reduces the number of measurement variables required. However, it does not easily account for temporal factors such as childhood experiences that may influence workaholism. In sum, learning theory provides generality (explains a large number of individual variances), parsimony (does not invoke extraneous variables), pragmatism (stimulates multiple hypotheses) and presents a feasible, if largely unexplored, basis for explaining workaholism (McMillan et al., 2001).

Trait Theory

Trait theory conceptualizes workaholism as a stable behavioural pattern that is dispositional (rather than environmental or biological), arises in late adolescence, stable across multiple workplaces, and is exacerbated by environmental stimuli (e.g., stress: McMillan et al., 2001). The parsimony of the theory, however, depends on whether trait-specific models or the more generic personality models are utilized.

Trait-specific models

Trait-specific models focus on narrow behavioural patterns and acknowledge individual variation (e.g., sibling differences), but explain a relatively restricted range of phenomena. For instance, obsessive compulsiveness explains task-focus but does address broader attitudes and values. The three most probable underlying traits in workaholism are obsessiveness, compul-

siveness, and high energy (Clark, Livesley, Schroeder, & Irish, 1996), each of which pertains to life in general rather than specifically to the work domain (McMillan et al., 2001). A broad range of data from well-validated measures support this theory of workaholism, especially with respect to obsessiveness, non-delegation, perfectionism, and hypomania (Clark et al., 1993; McMillan et al., 2002; Spence & Robbins, 1992). Obsessiveness has correlated with the drive component of workaholism at 0.35 ($r = 0.51$ when corrected for measurement error) and compulsiveness at 0.28 (0.37 corrected: McMillan et al., 2001). High energy levels (characteristic of hypomania) have related to workaholism at levels of 0.19, 0.25, & 0.27 (Clark et al., 1993; McMillan et al., 2001). This suggests that a combination of underlying traits may explain workaholism.

Generic personality models

Generic personality models explain more diffuse phenomena (e.g., conscientiousness), but sacrifice individual variability in the process. For instance, two people could be equally conscientious, but very different as people (McMillan et al., 2001). Clark et al. (1996) conceptualized workaholism as a pathological aspect of personality and found that workaholism related positively to the dimension of compulsiveness ($r = 0.41$) and to the higher order ('big five') trait of conscientiousness ($r = 0.53$). Thus, it is conceivable that workaholism is a lower order trait that relates in a hierarchical manner to higher order 'personality'.

Features, measures, and limitations of trait theory. While trait theory adopts a pessimistic view (workaholism is a part of personality and therefore relatively inflexible), it offers multiple explanations of workaholism. These range from simple characteristics like obsessiveness to broader aspects of the 'big five' personality factors, such as conscientiousness. Thus the theory is pragmatic, can be generalized, has broad utility, and is adequately supported by current data. Currently, there are two trait-based measures—SNAP-Work (Clark et al., 1993) and WorkBAT (Spence & Robbins, 1992)—and two relevant operational definitions. The first definition is an individual who is highly committed to work and devotes a good deal of time to it, which is evidenced by high involvemnt in work, compulsion to work and, low work enjoyment (Spence & Robbins, 1992). The second definition is a personal reluctance to disengage from work evidenced by the tendency to work (or to think about work) anytime and anywhere (McMillan et al., 2001). There are six boundaries and conditions required for trait theory to remain valid. These include: (i) the presence of environmental stimuli to trigger and maintain the behaviour, (ii) occurrence in some individuals within *all* societies, even during retirement, (iii) consistency across time, jobs, and life events, (iv) an inelastic tendency that can be modified to a slight degree but never entirely removed from a repertoire of behaviour, and (v) occurrence in the

presence of both positive and negative reinforcers. Patently, further development of the theory requires a longitudinal series of case studies, inter-organizational research and cross-cultural studies. Meanwhile, trait theory appears to provide a feasible basis from which to conduct further workaholism research.

Cognitive Theory

An important new development in workaholism research is the analysis of antecedent beliefs. This is based in cognitive theory, which proposes that people hold schemata (conceptual frameworks about the world) that are based in core beliefs, assumptions about causality, and automatic thoughts expressed as verbal self-statements (Beck, 1995). The theory predicts that workaholism arises from a core belief (e.g., I am a failure), consequent assumptions (e.g., if I work hard then I will not fail), and automatic thoughts (e.g., I must work hard). Thus, the beliefs, assumptions, and thoughts that activate workaholic behaviour become abbreviated over time to 'work equals worthiness,' and maintain high levels of workaholism. Burke (1999f), in an empirical investigation of the role of cognitions in workaholism, found that thoughts about striving against others, moral principles, and proving oneself predicted levels of workaholism. This holds important implications for workaholism, because if the data continue to support the theory, there are well-validated therapeutic interventions that modify such core beliefs (Beck, 1995). While it is premature to develop the theory further until further data emerge, this is a promising new development that warrants continued focus.

Family Systems Theory

A second, new, theoretical development arises from family-systems research. Family systems and, in particular, structural family theory consider that behaviour occurs in a context of interpersonal networks and dynamics, with a problem located within a system, as opposed to a person (Hayes, 1991). Thus, workaholism would be regarded as a family problem that arose from, and was maintained by, unhealthy dynamics. These dynamics may include blurred parent–child boundaries, over-responsibility, parentified children, circularity (everyone perpetuates the problem), enabling, concealment, and triangulation (parent–child alliances against the working partner: Robinson, 1998b, 2000b). For instance, an over-responsible person may express protectiveness for their family by overworking. The family, in turn, might enable the behaviour by cushioning the stress and hushing children when the worker arrives home. However, over time they may also perceive work as a tactic of distancing as opposed to protectiveness, and may respond by triangulating against the working partner. While these dynamics hold a small degree of face validity they make numerous assump-

tions and have not yet been subjected to empirical investigation. Clearly, before the theory can be further developed, we require empirical data with which to test its accuracy and appropriateness for workaholism.

Summary

While theoretical explication and development are still in their infancy, it is clear that trait theory has the foremost empirical support and learning theory provides the most convincing scientific utility (i.e., generality, parsimony, and pragmatism). Overall, therefore, a combination of trait and learning theories provides the most promising potential for future research and practical application. Thus, it appears that workaholism is currently most adequately explained as a personal trait that is activated and then maintained by environmental circumstances. From here on, therefore, it is imperative to instigate theoretically concordant research programmes that systematically test learning-trait hypotheses, accommodate them within empirically validated research designs, and apply them in practical settings. However, it is prudent to emphasize that the remaining theories may still provide valid explanations of the behaviour, but that their utility is constrained until more data are obtained.

RECENT RESEARCH

Given that workaholism research has generally progressed on an *ad hoc* basis, it is imperative that we start creating meaningful frameworks for summarizing and critiquing the increasing volume of data. The most parsimonious starting point is to use conventional psychological distinctions (e.g., antecedents, behaviour, and consequences) as a preliminary framework then move beyond those to more specific areas as the field matures. However, it is worth note that these three divisions are somewhat arbitrary and contain an implicit degree of overlap. Given this qualification, however, the present review will summarize the last five years' workaholism data in three sections: antecedents, workaholism behaviour, and consequences. First, it is first appropriate to provide brief comment on the general methodologies employed.

While the ensuing review covers all published articles available at the time of writing ($n = 34$), only 17 contain empirical data. Of these, all used questionnaire-based methodologies and yielded self-report data. Second, four employed WART (Robinson & Kelley, 1998, 1999; Robinson, Flowers, & Carroll, 2001; Robinson & Post, 1997), which still requires further validation (McMillan et al., 2001). It is also important to note that several arose from the same sample: Burke followed his initial publication (1999a) with nine further papers (cf., 2001b). Third, of the remaining studies, three employed Spence and Robbins's (1992)

WorkBAT (Bonebright, Clay, & Ankenmann, 2000; Kanai & Wakabayashi, 2001; Porter, 2001b), while the fourth (Mudrack & Naughton, 2001) involved a new measure of workaholism. Each of the following sections will present empirical data then a brief precis of the relevant publications.

Antecedents of Workaholism

As the majority of research has focused on describing, rather than explaining, workaholism, antecedents are currently the least understood aspect of workaholism, with only two published studies over the last five years. Before describing these studies, however, it is prudent to briefly describe the key components of workaholism and workaholic subtypes commonly referred to in the research. In general, as previously outlined, workaholism is believed to comprise up to three components, which include drive (an inner pressure to work), work enjoyment (work-related pleasure), and work involvement (psychological involvement with work in general: Spence & Robbins, 1992). [In the present context, the term 'work involvement' is workaholism-specific and not intended to be confused with the more traditional industrial psychology construct of work involvement. To retain consistency with other workaholism literature, the term will be used to refer *only* to Spence and Robbins' (1992) workaholism component in the remainder of the chapter.] As noted earlier, some researchers have argued that the work involvement component saturates the other two, and is therefore redundant (Kanai et al., 1996; McMillan et al., 2002). However, others have adopted Spence and Robbins's (1992) typology of workaholism, albeit based on work involvement scores. The three types are: work enthusiasts (high work involvement and enjoyment, low drive), non-enthusiastic workaholics (high work involvement, low enjoyment, high drive), and enthusiastic workaholics (high work involvement, enjoyment, and drive). Given the ongoing debate concerning the validity of the work involvement component, which is used to generate the types, the accuracy and validity of these subtypes remains unconfirmed. Thus, where research into the work involvement or workaholic subtypes is described in the upcoming sections, it is prudent to regard the findings as merely heuristic, rather than definitive, until further validation studies are published.

Empirical studies

In two of the first studies to concentrate on cognitive factors in workaholism, Burke (1999f, 2001b) investigated the predictive role of beliefs, fears, and perceptions. Burke proposed that there are two wellsprings of workaholism: individual differences (demographics, personality, family dynamics) and organizational characteristics (values that endorse work–personal life imbalance). The study utilized hierarchical regression analyses of scores on

WorkBAT (Spence & Robbins, 1992) with 530 MBA-qualified managers and professionals in Canada. Three groups of predictors were considered: (a) individual antecedents (beliefs and fears, perceived organizational support), (b) demographics (age, gender, and relationship status), and (c) work factors (seniority, size of organization, and tenure and role at the organization). The beliefs included striving against others, moral principles, and proving oneself, with each belief corresponding to a fear (e.g., I believe there can only be one winner in any situation: fear of failure). Personal demographics did not predict workaholism components and the workaholism–work involvement component was unable to be predicted by any variable. With respect to drive, work situation characteristics (seniority, less time in the current role) and cognitive antecedents (beliefs and fears, low perceptions of support for work–home balance) produced significant increments. With respect to enjoyment, work situation characteristics (predominantly seniority) and cognitive antecedents (weaker beliefs and fears, and higher perceptions of support for work–home balance) produced significant increments. However, these relationships were only of moderate strength. Finally, the three workaholic types (enthusiastic, non-enthusiastic, and work enthusiasts) had higher levels of Type-A-based beliefs and fears than non-workaholic subtypes (Burke, 1999f). Thus, cognitive antecedents appear to play a significant role in the development of the drive and enjoyment aspects of workaholism.

The second study to explore predictors of workaholism evaluated the role of job stressors in predicting workaholism. Kanai and Wakabayashi (2001) proposed that workaholism was a mode of adapting to a stressful work environment. They utilized hierarchical regression analyses of scores on the Japanese version of WorkBAT, which has two scales: joy and drive (Kanai et al., 1996). Participants were predominantly blue-collar Japanese males. Four groups of predictors were considered; (a) demographics (age, education, marital status, company size, change of job), (b) involvement variables (job time, job involvement, family time, family involvement), (c) job stressors (work overload quantity and quality, role conflict, and role ambiguity), and (d) work-related behaviours (perfectionism and non-delegation). Regression analyses showed that workaholism drive was predicted by demographics (company size, and marital status) job-time involvement, job involvement, family involvement, job stressors (work overload, role ambiguity), and work-related behaviours (perfectionism and non-delegation). Workaholism enjoyment was predicted by age, job-time involvement, job involvement, family involvement, family-time involvement, workload, and role ambiguity. The data supported the hypothesis that workaholism represents an attempt to adapt to job stressors, in particular quality and quantity of work overload. The study also gave some support for a lower prevalence of workaholism among blue-collar workers, although sampling biases (occupation, gender, education) may account for this finding.

Hypothetical speculations

In addition to the empirical data, several theorists have speculated about the antecedents of workaholism. Scott, Moore, and Miceli (1997) proposed that each different type of workaholism is associated with a different set of antecedents and suggested that researchers access the practitioner literature to generate hypotheses. Potential hypotheses include adrenaline addiction, addictive genetic predisposition, inadequate personal control, and learning 'opportunities' that strengthen underlying predispositions (Scott et al., 1997). Workaholism may also be linked with poverty, conflict at home (negative reinforcement), a voluntary phase of working a few extra hours (positive reinforcement), or an underlying trait (e.g., compulsiveness) activated in late adolescence (McMillan et al., 2001). Although several parcels of research have confirmed the trait-based links with workaholism, all used correlation statistics and failed to trace the relationship adequately to establish whether they were indeed antecedents, or rather, consequences of workaholism.

Behavioural Topography

Behavioural topography refers to the overt characteristics, structure, and magnitude of behaviour. For example, the most frequently cited topographical definition of workaholism is 'a desire to work long and hard (where) work habits almost always exceed the prescriptions of the job . . . and the expectations of the people with whom . . . they work' (Machlowitz, 1980, p. 11). However, while the topography of workaholism has been frequently discussed, this appears to be anecdotally rather than scientifically based, especially given the apparent lack of behaviour-observation studies. For instance, while early writers described workaholics as white-collar males who exhibited extremely poor balance between work and homes, and worked extremely long hours (Oates, 1968), there appears to have been no subsequent attempts to actually quantify the overt behaviour in an objective manner. However, studies are starting to emerge that evaluate the topography in at least a correlational manner.

Empirical studies

Empirical studies over the last five years are scant, but they focus on hitherto untested assumptions concerning gender, work–life balance, and stress.

Gender. The issue of gender differences in workaholism has been an interesting one; while the stereotype generally purports workaholics to be males, most studies have contradicted this (Burke, 2000b). Burke's study of Canadian managers described earlier investigated the relationship between particular workaholic behaviours and well-being within genders. Females reported higher levels of perfectionism and job stress that related

to lower levels of satisfaction and well-being, but were similar to males in terms of the three workaholism components (work involvement, drive, and enjoyment: Burke, 1999d). This study involved a relatively large sample size, virtually equal gender split, and a homogenous sample (across ethnicity, occupation, and education), which adds weight to the findings.

Work–life balance. The relationship between workaholism and the rest of life has been, until recently, the subject of much speculation, but little empirical testing. Bonebright et al. (2000) proposed that workaholism upsets the balance between work and personal time, and conducted one of the first empirical studies into the relationship between workaholism and work–life balance. They profiled predominantly male American employees into six work subtypes, using standardized scores and WorkBAT mean splits. Three groups of dependent variables were considered; (a) work–life conflict, (b) life satisfaction, and (c) purpose in life. Importantly, the data showed that the work involvement component related unpredictably to the first two variables ($r = 0.20, -0.200$) and had no relationship with life purpose. Drive demonstrated stronger trends while enjoyment had a non-significant relationship with work–life conflict, but related significantly to life satisfaction and life purpose. Workaholic subtypes had similar scores for work–life conflict, although enthusiastic workaholics had higher scores for life satisfaction and purpose in life. Although the authors argued that this provided evidence for continuing subtype distinctions in further research, the unpredictable performance of the work involvement factor (both here and in many previous studies) undermines this proposition.

Stress. The issue of stress in workaholism has been a continuing one. Porter (2001b) proposed that a work-addict is willing to sacrifice personal relationships to derive satisfaction from work. The study compared perfectionists with those who derive high joy from work across three groups of variables: (a) perceptions about organizational demands, (b) perception of risk-taking, and (c) beliefs about co-workers. In a sample of predominantly male, university-educated employees, Porter found no relationship between demographics or perceptions of organizational demands, with only enjoyment relating (negatively) to risk-taking. Those high in work enjoyment had consistently positive, team-focused beliefs about co-workers. These data provide some interesting challenges to the negative conception of enthusiastic workaholics that was implied by Spence and Robbins (1992).

Hypothetical speculations

While empirical data have been slow in evolving, hypothetical speculation has not. For instance, Robinson equated workaholics with ‘abusive workers’ and postulated that they differ from ‘healthy workers’ by the degree that work interferes with health, happiness, and relationships, as they lack the key attributes of optimal performers (warmth, outgoingness, and collaboration:

Robinson, 1997, 2000a). He has also claimed that workaholism involves 10 consistent patterns, three progressive 'stages', and 4 distinct subtypes (Robinson, 1996a, 1996b, 1997, 2000a). It is critical to emphasize that none of these propositions have been empirically tested. As Burke (2001a) and Robinson (2000a) both observed, we need a body of investigative studies using multiple techniques, and more persuasive validation data before these ideas can be substantiated.

However, in terms of providing a rigorous academic conceptualization, the Scott et al. (1997) work is invaluable. Based on a thorough review, comparison, contrast, and critique of the literature, they proposed three subtypes of workaholism. These are: (a) Compulsive Dependent (high stress; low job performance), (b) Perfectionist (high psychological problems; low job satisfaction), and (c) Achievement Oriented Workaholics (low stress; high creativity and performance). The authors provided an extensive theoretical analysis of the typology and inherent conceptual issues, but the construct and external validities of their model remain unexplored. Clearly, an empirical comparison of the Spence and Robbins (six) subtypes, Robinson (four) subtypes, and Scott et al. (four) subtypes is crucial to the satisfactory resolution of these differing hypothetical models.

Consequences of Workaholism

Empirical data

The consequences of workaholism have also been the focus of much conjecture, but limited scientific investigation. In general, however, the impact of workaholism is believed to extend to personal well-being and family satisfaction. The corresponding research from the last five years is addressed in detail below.

Well-being. The individual components of workaholism relate differently to psychological well-being and job stress (Burke, 2000c). In the study of Canadian managers, the work involvement component was unrelated to any of the measures. Workaholism drive related positively to psychosomatic symptoms and job stress but negatively to health-promoting behaviour and emotional well-being. Conversely, workaholism enjoyment related positively to health-promoting behaviour and emotional well-being and negatively to psychosomatic symptoms and job stress. Consequently, the three workaholic subtypes (Spence and Robbins, 1992) experienced differing levels of well-being. Burke (1999c) found that the three components consistently accounted for significant increments in explained variance on psychological well-being and even stronger amounts of variance in work outcomes and extra-work satisfactions. However, enjoyment and drive appeared to have the most influence, with enjoyment fostering satisfaction and well-being, and drive

yielding negative affect (Burke, 1999c). This trend was similarly evident in the female subset of the sample (Burke, 1999a).

Family impact. In one of the first studies to consider the effects of workaholism with children, Robinson and Kelley (1998) compared adult children of workaholics to adult children of non-workaholics. Using a family-systems paradigm, they proposed that workaholism was a harmful condition where workaholic parents created a home environment that increased the likelihood of poor psychological outcomes in their children. Four dependent variables were measured: depression, anxiety, self-concept, and external locus of control. University students were asked to estimate whether their parents were workaholics using WART. Self-identified children of workaholics had higher depression and external loci of control than others, but did not differ in respect of personal attributes or anxiety. Children of workaholic fathers indicated higher levels of anxiety, depression, and external locus of control than those of non-workaholic fathers. No differences were apparent for mothers. Importantly, the authors argued that the data patterns matched those of alcoholic families and implicated workaholic families as a 'diseased' family system where symptoms are passed onto children (Robinson & Kelley, 1998). While the study provides some fascinating hypotheses for prospective, longitudinal research, the present data are clearly limited to university-educated, female students who retrospectively perceive their fathers to be workaholic. A later study of students reported higher measures of depression and responsibility-seeking, and reported their parents worked longer hours than others (Carroll & Robinson, 2000). However, the use of third-party reports without collaborative data confounds the generality of these findings.

In the first study involving parents and young children, Robinson and Kelley (1999) measured workaholism in fourth- and fifth-grade children with WART, and found no relationship between parental workaholism and children's workaholism. Interestingly, teachers and children concurred on their ratings, while children's workaholism ratings related to anxiety, self-esteem, and locus of control. Thus it is possible that WART is actually tapping a broader construct, such as neuroticism or negative affect, and measurement issues confound the data. Other confounds include children's understanding of 'work' and the reliability with which they rate it. While this study is a valuable launch pad for future hypotheses, tighter methodologies are required to ascertain the true extent of the relationships between the variables.

Robinson and Post (1997) proposed that workaholism leads to poor family functioning and administered WART and a measure of family functioning to members of Workaholics Anonymous from workaholism conferences. Dependent variables included family problem-solving, communication, roles, affective responsiveness and involvement, behaviour control, and general functioning. Overall, the group of high-risk workaholics reported significantly worse functioning in almost every aspect than the low- and

medium-risk groups (Robinson & Post, 1997). A further investigation, using WART with female counselors revealed that workaholism had a negative impact on marital cohesion (Robinson et al., 2001). Outcome variables included marital disaffection (loss of emotional attachment, caring, and desire for emotional intimacy), positive feelings, and physical attraction. Data supported anecdotal observations of workaholism undermining marital stability. However, further testing is required to confirm the direction of this relationship, as marital cohesion may be affecting workaholism. In fact, Burke (1999b, 2000a) found that workaholism was unrelated to divorce, but did result in lower extra-work satisfactions (family, friends, community).

Hypothetical speculation

Hypothetical speculation about the consequences of workaholism has generally focused on family and organizational impact.

Family impact. Robinson (2001), in a review of family-systems–workaholism literature, cited negative interaction in family dynamics as a consequence of workaholism. He outlined the nature of these dynamics from a structural perspective (Robinson, 1998b), and hypothesized that spouses become extensions of the workaholics' ego, pseudo-single parents, and become aggressing partners in a pursuer–distancer dynamic (1998a). Specifically, the spouse may approach the worker for more intimacy, the worker may retreat as they already feel overloaded, the spouse makes a further approach (pursuit) and the worker makes another retreat (distancing), and thus the cycle perpetuates itself. Again, these hypotheses are scientifically untested, having arisen from anecdotal experience gained in counseling self-nominated workaholic families. Finally, in a broader systems analysis, Robinson (2000b) proposed that workaholics' spouses have ten characteristics. They feel: (a) ignored, (b) lonely, (c) second-rate, (d) subsumed to workaholics' demands, (e) controlled, (f) a need to seek attention, (g) their relationships are too serious, (h) guilty, (i) defective, and (j) uncertain about their sanity. Given that these descriptions appear somewhat pathologizing, they warrant immediate empirical attention, lest they become 'taken as fact' by the media and general public, without prior scientific verification.

Organizational impact. Porter (2001a) cautioned businesses not to confuse workaholism with high performance, as workaholism is destructive to both personal and professional relationships unless it is recognized and treated as an addictive behaviour. This built on an earlier conceptual review (Porter, 1996) that argued workaholism was evidenced within organizations by long working hours, high performance standards, job involvement, over-control, and personal identification with the job. Workaholics were purported to: choose solutions that were not congruent with the organizations' goals, sabotage efforts to promote work–home balance, have poor delegation, and overwork in the face of both failure and success (Porter, 1996). Burke (2000b)

recommended that employers support individual counseling, Workaholics Anonymous, workplace interventions (using performance and work habits as indicators of early problems), employee assistance programmes, reinforcement of incompatible alternatives (e.g., holidays), and workplace values that promote balanced priorities. In addition to the Scott et al. (1997) exploration of the consequence of workaholism subtypes, these papers provide interesting suggestions that are invaluable in generating further hypotheses.

FUTURE DIRECTIONS

The workaholism research arena is still in its infancy, with the extant research characterized by a strong focus on individual self-report data. Thus, there is ample opportunity to drive the field forward with innovative research designs and theoretically integrated research programmes. Accordingly, this section provides a brief critique of existing designs, before offering suggestions to guide future research.

Critique of Present Research Data and Designs

Overall, the current body of knowledge remains limited by the repertoire of methodologies employed, implicit value judgements, and the type of variables studied. As made apparent in the preceding review, the repertoire of research methodologies has been largely limited to questionnaire-based assessment of convergent constructs. Generally, research has relied upon self-report questionnaires and there is little information on how partners and work colleagues rate an individual's level of workaholism, how reliably existing measures perform against behavioural observations and physiological measures, or how workaholism changes over time. Unfortunately, the lack of validation research continues to restrict the utility of some of the promising models. It appears therefore that in the race to 'discover' new things about workaholism we have ignored a fundamental step: methodological diversity. While the resultant lack of creativity in research designs limits the current data, it also provides a substantial opportunity to enter the field and move things forward in a creative, novel, yet scientifically robust manner.

Unfortunately, despite Scott et al.'s (1997) caution, it appears that some researchers have not been restrained from making value judgements about workaholism, and there remains a continuing reluctance to investigate the possible positive outcomes of subtypes of workaholism. Thus, it is fair to critique much of the present research as biased in favour of pathological interpretation. Until we have data on the organizational value of workaholism (especially in terms of productivity, efficiency, and profitability) and the long-term outcomes of workaholism (using prospective designs) these conclusions remain premature. It is possible, for instance, that some

organizational cultures and job structures may be suited to workaholic types (Porter, 1996), or that positive aspects of workaholism could be trained into people (Scott et al., 1997). It is also important to question whether workaholism is necessarily bad for individuals. In this respect, in the race to 'adopt' addiction paradigms rather than critique their suitability, we have ignored another fundamental step: scientific neutrality. Again, substantial opportunities exist for new entrants to our field to think creatively and look critically at the value (or otherwise) of workaholism to all sectors of society, from employers and tax-takers to public-educators and health care-providers.

Finally, the range of correlates of workaholism that have been studied is relatively narrow. For example, given that workaholism appears to involve more time spent working than required, it is remiss that research has not addressed the impact of workaholism on 'outside of work' time (McMillan et al., 2001). We do not know, for instance, how much workaholic behaviour occurs outside the structured employment environment nor whether workaholics allocate less time to hygiene factors (e.g., diet and exercise). Additionally, the matching hypothesis (some spouses may not report relationship difficulties because both parties are in fact workaholic and thus compatible) appears to have been overlooked. It is arguable, therefore, that, in the race to 'capture' workaholism with pencil-and-paper self-report scales, we may have sacrificed ecological validity. It is therefore conceivable that alternate measurement methods (e.g., behavioural observation, third-party reports, triangulation) could more accurately capture the relationship between workaholism and a much broader range of constructs, such as organizational variables (e.g., productivity, citizenship: Burke, 2000b; Porter, 1996), cultural variables (e.g., race, ethnicity: Robinson, 2000b), and lifestyle variables (e.g., sexual orientation, childlessness, dual working status, socio-economic status).

Future Research Directions

The adoption of four new research designs would substantially benefit the growth of the field. These include: (a) contrasted groups, (b) alternating treatments, (c) longitudinal studies, and (d) heterogeneous sampling (McMillan et al., 2001). Contrasted group designs could elucidate how workaholic and non-workaholic behaviour differ in the workplace and in the community. This would enable a comparison of psycho-physiological data, clarify whether workaholics differ in terms of adrenaline levels, and determine whether addiction is indeed a key component of workaholism. Furthermore, these designs would allow researchers (and ultimately, employers) to differentiate between peak-performers, usual workers, and workaholics. As Robinson (2000b) noted, we also need ecological, systems-based designs to capture the subtleties involved in family dynamics. Additionally, investigation of the links between the home and work interfaces

would also be beneficial, so that the direction of the relationship can be quantified (Robinson, 2000b). Used strategically, therefore, contrasted designs would facilitate an *empirical* definition of the structural parameters of workaholism; a vital preliminary step in illuminating the relevance of addiction, learning, and trait theories (McMillan et al., 2001).

Alternating treatments would allow researchers to sequentially introduce independent variables (e.g., peer recognition, promotion) to ascertain how each impacts on workaholism. Baseline and alternating treatments (i.e., ABACAD) would enable functional analyses to determine functional equivalents and thus healthy substitutes for workaholism. Naturally this approach would involve complex ethical considerations (e.g., removing Treatment B in order to introduce Treatment C). However, this could be addressed, at least in part, by studying roles such as consulting, auditing, training, or self-employment, where an individual works from a 'home company' (A) but ventures into different clients' workplaces (B, C, and D) to undertake short-term projects. Workaholic symptoms could also be assessed under differing conditions (e.g., on holiday). Additionally, before–after case studies would elucidate the impact of changing jobs on an individual's workaholism, and permit comparison of workaholism levels *after* an influential workaholic has joined an organization. Overall, these designs would equip researchers to determine which variables modulate levels of workaholism. Used strategically, therefore, these designs could contribute substantially to our knowledge about the aetiological role of learning theory in workaholism.

There is also a clear need for longitudinal data. Many of the propositions from addiction theory could be clarified by longitudinal data that systematically eliminated plausible alternatives. Longitudinal designs could test the prediction of psychological addiction that workaholism is a progressive disease and evaluate the influence of developmental and life stressors (McMillan et al., 2001). Longitudinal data may explain the impact of different personality types on expression of workaholism. Finally, a sequential study of the antecedents, behaviour, and consequences of workaholism could shed light on the aetiological and maintaining factors of this little understood syndrome.

As outlined, homogeneous sampling (e.g., of purely degree-qualified professionals) has restricted the generality of much of the current data. Thus, the adoption of heterogeneous sampling strategies would contribute substantially to our knowledge about workaholism. Specifically, heterogeneous sampling of the general workforce (i.e., across all occupations, education levels, and income brackets) would facilitate an analysis of the organizational, cultural, and international prevalence of workaholism. Additionally, inter-organizational research could provide information about the influence of corporate culture and workaholic role models (Porter, 1996), while cross-sectional sampling would indicate whether some occupations

(e.g., entrepreneurs) have a higher incidence of workaholism. We could also benefit from information on prevalence rates in different strata of the population, such as the aged, and high earners, and cross-cultural comparisons of prevalence rates in different economies that elucidate whether workaholism is an individual or cultural variable. In sum, therefore, heterogeneous sampling is an absolutely imperative strategy for establishing the ecological validity of workaholism data.

SUMMARY

Workaholism occurs when a person has difficulty disengaging from work (evidenced by the capability to work at any time in any situation), a strong drive to work, and intense enjoyment of work. Most researchers concur that workaholism leads a person to work more hours, experience adverse health impacts, and to employ a differing use of leisure time than others. The recent surge in research interest and publications suggest that the dominant explanatory mechanisms include addiction, learning, trait, cognitive, and family-systems theories. Given the current breadth of empirical support, it appears that workaholism is most appropriately explained as a personal trait that is activated and maintained by environmental circumstances. While the majority of workaholism research has occurred on an *ad hoc* basis, confirmed antecedents include cognitions and job stress, while the behaviour itself appears to be gender-free, and the consequences include altered family dynamics and differing workplace behaviour. Hypothetical speculation includes antecedents of addiction, personality, learning and poverty, topographies consisting of up to six typologies, ten characteristics, and three stages, and consequences of poorer well-being, spousal and workplace relationships.

Overall, the resolution of the hypotheses outlined in the text requires innovative new research designs (e.g., contrasted groups, alternating treatments, longitudinal studies, heterogeneous sampling, to name a few). Additionally, as the world trend toward globalization, international migration, cross-cultural and electronic communications, multinational production lines, mobilized technology, and elastic-boundaried workplaces (e.g., working from home, work, abroad) shrink the distance between workplaces, homes, cultures, and countries, the need for workaholism research will increase. Along with it, so will the opportunities for critical-thinking scientists with an innovative and broad methodological approach to enter the workaholism research arena. Thus, while workaholism research has made some vital progress over the last five years, the future holds as numerous opportunities as the current literature provides hypotheses.

ACKNOWLEDGMENTS

Lynley McMillan was supported by a Foundation for Research Science and Technology Bright Futures Fellowship while writing this review. Ronald Burke was supported by the School of Business, York University, Canada.

REFERENCES

Beck, J. S. (1995). *Cognitive Therapy: Basics and Beyond*. New York: Guilford Press.

Bonebright, C. A., Clay, D. L., & Ankenmann, R. D. (2000). The relationship of workaholism with work-life conflict, life satisfaction, and purpose in life. *Journal of Counseling Psychology*, **47**(4), 469–477.

Burke, R. J. (1999a). Workaholism among women managers: Work and life satisfactions and psychological well-being. *Equal Opportunities International*, **18**(7), 25–35.

Burke, R. J. (1999b). Workaholism and extra-work satisfactions. *International Journal of Organizational Analysis*, **7**(4), 352–364.

Burke, R. J. (1999c). It's not how hard you work but how you work hard: Evaluating workaholism components. *International Journal of Stress Management*, **6**(4), 225–239.

Burke, R. J. (1999d). Workaholism in organizations: Gender differences. Sex Roles, 41, (5–6), 333–345.

Burke, R. J. (1999e). Workaholism in organizations: Measurement validation and replication. *International Journal of Stress Management*, *6*(1), 45–55.

Burke, R. J. (1999f). Workaholism in organizations: The role of personal beliefs and fears. *Anxiety, Stress and Coping*, **00**, 1–12.

Burke, R. J. (2000a). Workaholism and divorce. *Psychological Reports*, **86**(1), 219–220.

Burke, R. J. (2000b). Workaholism in organizations: Concepts, results and future directions. *International Journal of Management Reviews*, **2**(1), 1–16.

Burke, R. J. (2000c). Workaholism in organizations: Psychological and physical well-being consequences. *Stress Medicine*, **16**, 11–16.

Burke, R. J. (2001a). Editorial: Workaholism in organizations. *International Journal of Stress Management*, **8**(2), 65–68.

Burke, R. J. (2001b). Predictors of workaholism component and behaviors. *International Journal of Stress Management*, **8**(2), 113–127.

Carroll, J. J., & Robinson, B. E. (2000). Depression and parentification among adults as related to parental workaholism and alcoholism. *Family Journal of Counseling and Therapy for Couples and Families*, **8**(4), 360–367.

Clark, L. A., Livesley, W. J., Schroeder, M. L., & Irish, S. L. (1996). Convergence of two systems for assessing specific traits of personality disorder. *Psychological Assessment*, **8**, 294–303.

Clark, L. A., McEwen, J. L., Collard, L. M., & Hickok, L. G. (1993). Symptoms and traits of personality disorder: Two new methods in their assessment. *Psychological Assessment*, **5**, 81–91.

Cross, G. (1990). *A Social History of Leisure Since 1600*. State College, PA: Venture.

Di Chiara, G. (1995). The role of dopamine in drug abuse viewed from the perspective of its role in motivation. *Drug and Alcohol Dependence*, **38**, 95–137.

Eysenck, H. J. (1997). Addiction, personality and motivation. *Human Psychopharmacology*, **12**, S79–S87.

Fassel, D. (1992). *Working Ourselves to Death*. London: HarperCollins.

Hayes, H. (1991). A re-introduction to family therapy: Clarification of three schools. *Australia and New Zealand Journal of Family Therapy*, **12**(1), 27–45.

Kanai, A., & Wakabayashi, M. (2001). Workaholism among Japanese blue collar employees. *International Journal of Stress Management*, **8**(2), 129–145.

Kanai, A., Wakabayashi, M., & Fling, S. (1996). Workaholism among employees in Japanese corporations: An examination based on the Japanese version of the Workaholism Scales. *Japanese Psychological Research*, **38**(4), 192–203.

Machlowitz, M. M. (1980). *Workaholics: Living with Them, Working with Them*. Reading, MA: Addison-Wesley.

McMillan, L. H. W., Brady, E. C., O'Driscoll, M. P., & Marsh, N. V. (2002). A multifaceted validation study of Spence and Robbins' (1992) Workaholism Battery. *Journal of Occupational and Organisational Psychology*, **75**(3), 357–368.

McMillan, L. H. W., O'Driscoll, M. P., Marsh, N. V., & Brady, E. C. (2001). Understanding workaholism: Data synthesis, theoretical critique and future design strategies. *International Journal of Stress Management*, **8**(2), 69–91.

Mudrack, P. E., & Naughton, T. J. (2001). The assessment of workaholism as behavioral tendencies: Scale development and preliminary empirical testing. *International Journal of Stress Management*, **8**(2), 93–111.

Oates, W. E. (1968). On being a 'workaholic' (A serious jest). *Pastoral Psychology*, **19**, 16–20.

Porter, G. (1996). Organizational impact of workaholism: Suggestions for researching the negative outcomes of excessive work. *Journal of Occupational Health Psychology*, **1**, 70–84.

Porter, G. (2001a). Workaholics as high performance employees: The intersection of workplace and family relationship problems. In B. E. Robinson, & N. D. Chase (eds), *High Performing Families: Causes, Consequences and Clinical Solutions* (pp. 43–69). Alexandria, VA: American Counseling Association.

Porter, G. (2001b). Workaholic tendencies and the high potential for stress among co-workers. *International Journal of Stress Management*, **8**(2), 147–164.

Preece, W. E. (ed.). (1981). *Encyclopaedia Britannica* (15th edn, Vol. 19, pp. 932–942). Chicago: Encyclopaedia Britannica.

Robinson, B. E. (1996a). The psychosocial and familial aspect of work addiction: Preliminary perspectives and hypotheses. *Journal of Counseling and Development*, **74**, 447–452.

Robinson, B. E. (1996b). Relationship between work addiction and family functioning: Clinical implications for marriage and family therapists. *Journal of Family Psychotherapy*, **7**(3), 13–29.

Robinson, B. E. (1997). Work addiction: Implications for EAP counseling and research. *Employee Assistance Quarterly*, **12**(4), 1–13.

Robinson, B. E. (1998a). Spouses of workaholics: Clinical implications for psychotherapy. *Psychotherapy*, **35**(2), 260–368.

Robinson, B. E. (1998b). The workaholic family: A clinical perspective. *The American Journal of Family Therapy*, **22**, 65–75.

Robinson, B. E. (1998c). *Chained to the Desk: A Guidebook for Workaholics, Their Partners and Children, and the Clinicians who Treat Them*. New York: New York University Press.

Robinson, B. E. (2000a). A typology of workaholics with implications for counselors. *Journal of Addictions and Offender Counseling*, **21**, 34–48.
Robinson, B. E. (2000b).Workaholism: Bridging the gap between workplace, sociocultural, and family research. *Journal of Employment Counseling*, **37**, 31–47.
Robinson, B. E. (2001). Workaholism and family functioning: A psychological profile of family members. In B. E. Robinson, & N. D. Chase (eds), *High Performing Families: Causes, Consequences and Clinical Solutions* (pp. 3–22). Alexandria, VA: American Counseling Association.
Robinson, B. E., Flowers, C., & Carroll, J. (2001). Work stress and marriage: A theoretical model examining the relationships between workaholism and marital cohesion. *International Journal of Stress Management*, **8**(2), 165–175.
Robinson, B. E., & Kelley, L. (1998). Adult children of workaholics: Self-concept, anxiety, depression, and locus of control. *The American Journal of Family Therapy*, **26**, 223–238.
Robinson, B. E., & Kelley, L. (1999). School age workaholic children: Type A behaviors, self esteem, anxiety and locus of control. *Early Child Development and Care*, **158**, 43–50.
Robinson, B. E., & Post, P. (1997). Risk of addiction to work and family functioning. *Psychological Reports*, **81**, 91–95.
Robinson, J. P., & Godbey, G. G. (1992). *Time for Life: The Surprising Ways Americans Use Their Time*. University Park, PA: Pennsylvania State University Press.
Rohrlich, J. (1980). *Work and Love: The Crucial Balance*. New York: Harmony.
Scott, K. S., Moore, K. S., & Miceli, M. P. (1997). An exploration of the meaning and consequences of workaholism. *Human Relations*, **50**, 287–314.
Skinner, B. F. (1974). *About Behaviorism*. London: Penguin.
Spence, J. T., & Robbins, A. S. (1992). Workaholism: Definition, measurement, and preliminary results. *Journal of Personality Assessment*, **58**, 160–178.

Chapter 6

ETHNIC GROUP DIFFERENCES AND MEASURING COGNITIVE ABILITY

Helen Baron[1] and Tamsin Martin
SHL Group plc

and Ashley Proud, Kirsty Weston, and Chris Elshaw
QinetiQ Ltd

In the domain of occupational psychology, two of the most consistent findings relate to cognitive ability test scores: they are highly predictive of job performance and they result in substantial score differences between racial and ethnic groups. In use, score differences lead to adverse impact in decision-making where disproportionately more members of the lower scoring group are excluded.

This chapter reviews current findings on ethnic group differences. While the vast majority of published research is of US origin, we take a more international perspective and consider the available findings from elsewhere in the world. We look at the evidence of when and where differences occur, and the factors that impact on their magnitude, on differential validity, and on decision-making processes. We review the evidence regarding various explanations of group differences that have been suggested and the latest practical approaches being put forward to try to reduce adverse impact in practice.

GROUP DIFFERENCES

The group differences literature relates to a wide variety of measures including IQ tests, general ability batteries with high cognitive loadings and tests of specific skills with lower cognitive loadings. As well as the use of tests in an

[1] Now an independent consultant.

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

occupational context, similar measures are used in educational and military contexts, as well as for research. Most of the literature on race group differences comes from the USA with the main focus on comparisons between Blacks and Whites. With this wealth of factors potentially affecting results, it is important not to generalize findings to other types of measures, other contexts or other groups without justification. We present findings from the USA first, followed by others from around the world, and present the differences in terms of the pooled, within-group standard deviation (SD), often called the *d* statistic.

United States

Black–White comparisons

The generally accepted figure for Black–White group differences in the USA is one standard deviation (SD) with the White group scoring higher. However, recent evidence suggests that the situation is more complex and that this overall figure is more variable than first thought. Many studies are reliant on a small number of large data sets. The use of the Wonderlic Test and General Aptitude Test Battery (GATB) dominates the reported studies. These large data sets may have an undue influence on the generalizability of many results. Roth, Bevier, Bobko, Switzer, and Tyler (2001) excluded these data sets from their meta-analysis, encompassing data from both employment and educational settings, and found the overall uncorrected difference between Whites and Blacks to be 1.10 SD ($k = 105$), a little above the generally accepted 1.0 SD difference. Occupational samples showed differences 0.1 SD lower than military and educational samples.

In addition to some specific, large data sets, several other sampling factors, such as job complexity and employment status, have been shown to have moderating effects on observed differences.

Job complexity. Within occupational samples, Roth et al. (2001) found that job complexity was a strong moderator of group differences. Splitting jobs into low, moderate, and high complexity, they found the smallest difference (0.63 SD) for the most complex jobs and the largest (0.86 SD) for the least complex jobs. These differences could be explained through a degree of self-selection. Individuals apply for jobs they perceive as appropriate to their ability and therefore those with lower ability are less likely to apply for a job with high-level cognitive demands and vice versa. The self-selection hypothesis is also supported by reduced differences in studies of single jobs (0.74 SD for applicants and 0.38 SD for incumbents) compared with across-job studies. Within occupational studies, employment status also acts as a moderator of group differences. Applicant differences in general cognitive ability in occupational samples averaged 0.99 SD, whereas the incumbent differences were 0.41 SD.

Non-occupational samples. Military samples (Roth et al., 2001) showed a more extreme pattern of differences with Whites scoring 1.46 SD higher than Black applicants. Differences among incumbents were much smaller, reducing to 0.53 SD.

Educational samples also show variability in group differences, although not the same pattern. Roth et al. (2001) report a difference of 0.95 SD and 0.98 SD for high-school samples and college applicants on the Scholastic Assessment Test (SAT) and the American College Test (ACT) data, respectively. Lynn (1996) found a difference of 13.46 IQ points between Blacks and Whites on general cognitive ability for high-school children (6–17-year-olds).

Roth et al. (2001) found that college-students showed a smaller difference of 0.69 SD between Blacks and Whites. The size of the college-student group difference may be a function of the within-school analysis and the fact that college populations are pre-selected. The difference between scores for Black and White graduate-school applicants was found to be almost double at 1.34 SD.

Camara and Schmidt (1999) report a similar pattern on college and post-graduate entrance examinations. Differences ranging from 0.82 SD to 0.98 SD were found on college entrance examinations (the SAT and the ACT) and from 0.96 SD to 1.14 SD for graduate entrance examinations (e.g., the Graduate Record Examination, GRE, the Law School Admission Test, LSAT, or the Medical College Admissions Test, MCAT).

Type of ability. Schmitt, Clause, and Pulakos (1996) found differences between Black and White groups varying from 0.83 SD for general cognitive ability to 0.14 SD for manual dexterity in their meta-analysis. Group differences on spatial, verbal, and mathematical ability were all around 0.6 SD. However, Loehlin, Lindzey, and Spuhler (1975) found the largest differences between the groups on spatial ability. Lynn (1996) found the least differences on verbal tests, with the largest differences on spatial tests. Verive and McDaniel (1996) studied short-term memory tests of various kinds and reported smaller group differences than for general cognitive ability. They found a difference of 0.48 SD between Whites and pooled ethnic minorities in their meta-analysis of applicant data ($N = 27{,}793$).

Within occupational samples, Roth et al. (2001) found that, with GATB data excluded, verbal and mathematics scores showed very similar group differences of 0.76 SD, and 0.71 SD respectively. In educational samples, Roth et al. (2001) found the largest differences in total scores (0.97–1.34 SD), with the smallest differences in the ACT and GRE tests on mathematics tests (0.82–1.08 SD), but not in the SAT tests where verbal scores showed the smallest difference (0.84 SD).

Hough, Oswald, and Ployhart (2001) review a broad span of studies and suggest effect sizes of 1.0 SD for general cognitive ability, 0.7 SD for quantitative and spatial ability, 0.6 SD for verbal ability, 0.5 SD for memory, and 0.3 SD for mental processing speed.

Hispanic–White comparisons

Compared with the number of studies reporting Black–White group differences, there are few studies including a Hispanic group. Available findings typically report Hispanics' scores lower than White groups but a little higher than Black groups. Overall differences between Whites and Hispanics range from 0.5 SD (Gottfredsen, 1988; Hough et al., 2000; Schmitt et al., 1996) to 0.8 SD (Sackett and Wilk, 1994). Lynn (1996) found Hispanics to have a mean of 8.79 IQ points less than Whites. Roth et al. (2001) in their meta-analysis found an overall difference of 0.72 SD across all samples, but somewhat larger differences were observed within the occupational samples (0.83 SD). However, the largest differences are seen with the data sets using the Wonderlic Test. When this test was excluded, the difference for occupational samples decreased to 0.58 SD.

Non-occupational samples. In the educational sphere, Camara and Schmidt (1999) report differences between Hispanics and Whites of 0.5 SD to 0.63 SD favouring the White group at college entrance. However, more variability was found at the postgraduate level with differences ranging from 0.46 SD to 1.0 SD. Military samples showed similar effect sizes (0.85 SD) to occupational ones in the Roth et al. (2001) study.

Type of ability. Studies that reported scores for specific ability tests also revealed a range of values. Overall, the differences seemed to be smallest for numerical and mathematical tests and largest for tests of verbal ability. Hough et al. (2001) report a general cognitive ability difference of 0.5 SD, 0.4 SD for verbal ability and mental processing speed, and 0.3 SD for quantitative tests. Lynn (1996) found greater differences on verbal than spatial ability. Roth et al. (2001) found a d of 0.4 for verbal tests and 0.28 for quantitative tests in occupational samples. Lower verbal reasoning scores may be related to language skills (see Roth et al., 2001). Data from the 1980 Census Public Use Sample shows that, at that time, one-quarter of Puerto Rican and Mexican Americans and over two-fifths of Cubans speak English 'not well' or 'not at all' (Rodriguez, 1992). Clearly, verbal tests are most likely to be affected by poor command of English.

Asian American (Far Eastern)–White comparisons

There is substantially less literature and research on comparisons between Asians and Whites. However, what there is suggests that Asians often outperform White groups in many domains (Neisser et al., 1996). The differences reported are of a much smaller magnitude so have less impact on employment or educational opportunities than Black–White differences (Roth et al., 2001).

Lynn (1996) found that Asians (of high-school age) scored a mean of 4.42

IQ points higher than Whites. Hough et al. (2001) estimate a difference favouring Asians of 0.2 SD.

Non-occupational samples. In an educational setting, Camara and Schmidt (1999) reported differences ranging from scores of 0.29 SD lower than Whites through to 0.46 SD higher than Whites, on a number of different college entrance tests.

Type of ability. Camara and Schmidt (1999) found that typically, the Asian group outperformed Whites on quantitative tasks but had a similar level of performance on verbal measures. Lynn (1996) found the largest Asian–White differences on non-verbal reasoning and spatial ability.

Around the World

There is far less published literature on group differences in other countries. The following trends are based on both published and unpublished findings from a variety of sources. The nature of the tests, the contexts of measurement, and the group comparisons of interest all differ from country to country. However, there is a consistent picture of lower performance on cognitive ability tests among socially disadvantaged groups.

Canada

Chung-Yan, Hausdorf, and Cronshaw (2000) found differences of about 0.75 SD on urban transit applicants using the GATB in Canada, with the majority White group scoring higher than the minority group. The minority group was mixed, including Blacks and people from various parts of Asia as well as those from First Nations (native Americans).

United Kingdom

The most recent estimates for the UK suggest that 93% of the population is White with some 7% from ethnic minority groups (Office of National Statistics, 2001). Of these around two-thirds are of Asian origin and one-third is Black, the majority originating in the Caribbean. There are a very small number of Asians of Chinese or other Far-Eastern origin; the vast majority are from the Indian subcontinent with origins in Pakistan, Bangladesh, and India. Although there have been Black people in Britain for hundreds of years, the majority of these populations are the result of immigration from former British colonies during the second half of the 20th century.

Military studies provide information on group differences in the UK on tests of general cognitive ability. Cook (1999a) provides a comprehensive examination of the adverse impact and differential validity of the British Army Recruit Battery test (BARB), which is taken by all non-officer entrants

into the British Army. The BARB comprises six speeded, computer-administered, item-generative tests, which together are considered to provide a measure of *g*, or general cognitive ability. The tests involve multiple-choice responses to relatively simple matching or contrasting item content and place high demands on working memory. In this study, Whites (N = 31,947) typically outperformed a pooled ethnic minority group (N = 581) by around 0.4 SD.

Cook (1999b) also examined the Royal Navy's Recruiting Test (RT), which is another test of cognitive ability administered to all non-officer applicants. The RT comprises a battery of four more traditional tests, including numeracy, literacy, and mechanical comprehension. Overall, Whites (N = 20,891) scored around 0.5 SD higher than a pooled ethnic minority group (N = 254). A mixed sample of school students and applicants showed larger differences of 0.8 SD (Mains-Smith & Abram, 2000).

Most of the available UK data contrast the White group with a pooled group of ethnic minority candidates because the base rate for the different ethnic groups is low. Some further analysis of the ethnic groups was possible in Cook's studies (1999a and 1999b)—see Table 6.1.

Type of ability. Table 6.2 shows results from available data for a range of different tests of specific abilities at varying levels designed for use in occupational and military settings. The vast majority of the data are from applicant groups although some are from incumbents or student trial samples. It includes the studies described above as well as data collected by test publishers. The samples are of varying sizes with a total sample size of several tens of thousands.

Table 6.2 shows Whites consistently scoring higher than the pooled ethnic minority sample. Overall group differences range from 0.16 SD to 1.09 SD. The few cases where separate data were available for Asian and Black groups suggest that both have lower average scores than the White group, with Asians scoring slightly higher than Blacks in general. However, there is

Table 6.1 UK group difference (*d*) results from military tests. Negative *d* values show ethnic minority groups scoring higher than the White group.

	RT (Navy)	*N*	BARB (Army)	*N*
Black (African)	0.3	25	0.7	65
Black (Caribbean)	0.5	54	0.4	170
Black (other)	0.5	75	0.4	175
Indian	0.4	36	0.1	58
Pakistani	0.6	43	0.6	84
Bangladeshi	1.3	6	1.2	12
Chinese	– 0.07	15	– 0.5	17
Total		254		581

Table 6.2 Sample weighted effect sizes for various samples and tests: White–pooled ethnic minority comparisons. All *d* values show Whites scoring higher than the pooled ethnic minority groups.

Type of test and sample	General ability	Verbal	Numerical	Diagrammatic Spatial Mechanical Technical	Clerical
Military					
BARB (Cook, 1999a)	0.40				
N (White/ethnic minority)	38,959/581				
RT students and applicants (Mains-Smith & Abram, 2000)	0.80	0.75	0.50	0.89	
N (White/ethnic minority)	779/554				
RT applicants (Cook, 1999b)	0.47	0.57	0.23	1.09	
N (White/ethnic minority)	20,773/254				
Occupational					
SHL Group (2002) Low-complexity jobs		0.45	0.38		0.41
N (White/ethnic minority)		2,951/1,153	2,549/870		3,941/1,177
SHL Group (2002) Moderate-complexity jobs		0.63	0.53	0.31	0.73
N (White/ethnic minority)		3,578/518	3,798/512	961/487	1,819/148
SHL Group (2002) High-complexity jobs		0.30	0.60	0.64	
		1,221/159	1,129/140	262/64	
ABLE—mixed skill requirements (Deakin, 2000)			0.16–0.50		
N (White/ethnic minority)			1,685/274		

considerable variation within Asian groups, with those of Indian and Chinese origins performing particularly well, and those of Bangladeshi origin performing relatively poorly, as seen in the military data of Table 6.1. In contrast with typical US findings, there is a great deal of variance between the different samples. This may be partly due to sampling error but Baron and Chudleigh (1999) report a series of samples from one organization where the trend is reversed with Whites performing less well than the pooled ethnic minority group, suggesting that there could be more systematic variation (e.g., through self-selection for jobs).

It has been suggested that the language barrier may be part of the reason behind the racial group differences in cognitive testing. Cook (1999a) found, in his study of the BARB, that for the majority of the candidates English was the language they used most often. Only the Bangladeshi group had a high proportion (16.7%) of people for whom English was not the main language. This group scored 1.2 SD below Whites and was the lowest scoring group of all. These results lend support to the assertions of Rodriguez (1992) and Roth et al. (2001) that weaker language skills affect performance, even on tests of general cognitive ability.

Overall, typical differences seem to be around 0.5 SD, although this is masking variability for job complexity and language issues for some groups. In addition the use of pooled data for various ethnic minority groups may well be hiding different patterns for the different groups.

Netherlands

A study in the Netherlands (te Nijenhuis & van der Flier, 1997) compared Dutch language GATB scores among applicants to the railways for first generation immigrants ($n = 1{,}322$) and a matched group from the majority population ($n = 806$). The largest groups of immigrants came from Surinam (40%)—mainly Creoles and Asian Indians—Turkey (21%), North Africa (13%), and Dutch Antilles (9.5%)—mainly Creoles. Differences consistently favored the majority group and ranged from 0.1 SD for the Mark Making test (a measure of Aiming) to 1.5 SD for the Vocabulary test. The Arithmetic Reasoning and Three-Dimensional Space tests also showed differences above 1. Results for the different groups were generally similar, but the North African group consistently showed the lowest scores. The authors suggest that language is an important factor for a large proportion of the immigrant group.

Singapore

Two samples of applicant data were available from Singapore, relating to a variety of different tests for jobs at varying levels of complexity. The samples were predominantly mixed groups. The tests were administered in English in accordance with local practice. English is the accepted business language and its use is emphasized in schools. One sample consisted predominantly of Singaporean Chinese with about 20% Singaporean Indian, Malays, and Eurasians ($N = 640$). The second sample also included 30% Malays and 20% Thai, Vietnamese, and Indonesian ($N = 115$). These samples were compared with equivalent UK data from the same tests. The Singaporean samples performed consistently better on all tests than the UK ethnic minority groups (0.16–1.1 SD higher), but scored lower on all verbal tests than the UK White groups by between 0.2 SD and 0.7 SD. However, the Singaporean sample outperformed the UK White group on numerical tests by 0.2 to 0.3 SD. This supports the suggestion that differences on verbal tests are related to language skills, as many of the candidates took the tests in a second or third language (SHL Group, 2002).

China

A number of samples of applicant data from Hong Kong, Taiwan, Macau, and Mainland China was available (SHL Group, 2002). Again, these results

refer to English language test administration reflecting common local practice. The results show that on the verbal tests the Chinese sample (*N* = 11,957) performed 0.5 to 0.7 SD below equivalent UK pooled ethnic minority samples and 1.1 SD below UK White groups on the same tests. The differences were less marked on numerical tests, with the Chinese (*N* = 10,496) outperforming the UK ethnic minority group in all cases. For moderate-complexity jobs, the Chinese group score 0.2 SD below the equivalent UK White group. For graduate-level jobs the Chinese outperform UK Whites on numerical tasks by 0.2 SD—a similar difference to that found in US and Singapore data.

Within the total China sample itself, there are some interesting and significant differences between the four regions (Hong Kong, Taiwan, Macau, and Mainland China). Hong Kong managers and graduates outperformed a pooled sample from all other regions in both verbal (Hong Kong *N* = 4,747, pooled sample *N* = 5,699, 0.68 SD) and numerical (Hong Kong *N* = 4,747, pooled sample *N* = 203, 1 SD) tests. The differences between Hong Kong and the other regions on the verbal tests are smaller than the differences shown on the numerical tests. However, the size of the pooled sample completing numerical tests is small.

New Zealand

Researchers in New Zealand have in the past reported 'a deep mistrust by New Zealanders of tests of ability and aptitude' (McLellan, Inkson, Dakin, Dewew and Elkin, 1987, p. 80). However, the use of ability and aptitude tests has become more widespread in recent years, although the relevant research examining the effects of these tests on New Zealand's different ethnic groups remains sparse. Some data contrast findings from the White majority with those for the Maori group. Only very small samples were available, but differences were typically in favour of Whites of the order of 0.5 SD (SHL Group, 2002). Flynn (1988; in Salmon, 1990) reported a performance gap between Maori and White New Zealanders on tests of cognitive ability that was not reduced by the use of non-verbal scales.

South Africa

In data from South Africa, Whites are very much the minority group. However, in socio-economic terms they have a very great advantage over Black, Asian, and Coloured groups, and generally greatly enhanced educational opportunities. There are 11 official languages in South Africa, with all students being required to study at least two. Students may choose to study in Afrikaans and a Black language—this will put them at a disadvantage when taking cognitive ability tests in English, on which these data are based.

English language testing is common practice where English fluency levels are high. For Afrikaner Whites, English is also a second language.

Data are available for verbal, numerical, diagrammatic, clerical, and spatial tests, mainly with applicants to jobs at a variety of levels and industries but also some students and applicants for IT courses. Altogether, the sample consists of over 9,000 Whites and 9,000 Blacks, Asians, and Coloureds. Whites consistently scored higher than the combined sample of Blacks, Asians and Coloureds. Differences ranged from 0.62 SD on clerical tests and 0.66 SD on numerical tests, to 0.72 SD on verbal tests and 1.27 SD on diagrammatic tests (SHL Group, 2002).

Nigeria

A multi-national employer tested a large number of local Nigerians ($N > 2{,}300$) for employment across a range of different tests (including verbal, numerical, diagrammatic, and spatial), reflecting several different jobs. The sample was almost entirely Black. Testing was in UK English, which was generally not the first language of candidates but was often the only language in which they were literate. Results showed the Nigerian group scored between 0.63 and 1.55 SD lower than UK pooled ethnic minority groups and between 0.78 and 1.79 SD lower than a UK White group on the same tests (SHL Group, 2002). These group differences could be the result of a number of different factors, including unfamiliarity with testing, poor educational opportunities, and language issues.

Summary

The latest US findings show that, while average scores for Whites are consistently higher than Blacks and Hispanics across the whole range of the cognitive ability domain, the size of the differences observed is moderated by a number of factors including the type of skill, job complexity or educational level, and study design.

There are consistent findings of score differences between groups in countries around the world. Actual groups and differences vary, but, in general, groups that have lower socio-economic status and poorer educational opportunities show lower test scores on the majority of tests than more privileged groups. These groups also typically come from cultural traditions that are very different from Western culture, from which cognitive ability testing approaches spring.

The international findings, while much less comprehensive, do seem to mirror US findings in some respects, with Black groups performing less well than Whites, and Asians from the Far East performing better than Whites. For groups which are likely to have a poorer command of English, the largest differences are often seen on verbal tasks. Otherwise the largest

differences seem to occur on the general measures in the US data and with groups who have the least access to education or knowledge of Western culture.

Roth et al. (2001) suggest that differences tend to increase with the saturation of *g* in the measure of ability. This hypothesis has sometimes been called 'Spearman's hypothesis'. If there is an underlying difference in *g* between two groups, then this will be reflected more strongly in measures with a stronger *g* loading. Nyborg and Jensen (2000) argue strongly in support of the hypothesis although others (e.g., Schoenemann, 1997; Kadlec, 1997) have argued that their findings could be due to statistical artefacts.

VALIDITY

The main use of cognitive ability tests in industrial and organizational psychology is in predicting job performance. This underpins selection and promotion applications as well as use in more developmental contexts such as career counselling. There is strong evidence to support their usage in these contexts. In the first meta-analysis of the relationship between job performance and cognitive ability tests, Hunter and Hunter (1984) found an average validity of 0.53, after correction, for cognitive ability tests when predicting performance in entry-level jobs. Wise, McHenry, and Campbell (1990) found significant correlations between cognitive ability test scores and performance in the large-scale US Army Project A study. Schmidt and Hunter's (1998) review of validity findings supported their original result. These findings are based almost entirely on US data. There is much less published data from the rest of the world, but where studies have been done, findings have been similar. Robertson and Kinder (1993) found an average uncorrected validity for cognitive ability tests in the UK with managerial groups of 0.02–0.36 for a range of criteria. Nyfield, Gibbons, Baron, and Robertson (1995) found similar validities for cognitive ability tests for UK, US, and Turkish samples in a concurrent study of international managers. Salgado and Anderson (2002) found uncorrected validities of 0.36 and 0.18 against job performance ratings in a meta-analysis of Spanish and UK studies, respectively.

Differential Validity

Overall validity could mask a situation where the test was predictive of performance for only one group, or was a better or different predictor for one group than the other. This could mean that the difference in scores observed for different racial groups was not reflected in similar differences in performance and therefore would result in unfair selection decisions. Hunter, Schmidt, and Hunter (1979), in a meta-analysis of 39 studies where Black

and White validities were reported separately, found very similar results for the two groups. Humphreys (1992) suggests that IQ tests are equally predictive of Black and White criterion performance. Schmidt, Pearlman, and Hunter (1980) found no differences in validity for Hispanic and White groups. Hartigan and Wigdor (1989) found similar validities for Black and White groups in their study of GATB findings, although there was a marked trend for the Black group validities to be lower. Jones and Raju (2000) find a significant difference in regression lines for Black and White applicants to an apprentice training scheme using vocabulary test scores to predict first-year Grade Point Average (GPA) on the training programme. The test was a better predictor for the White group, and the common regression line tended to overpredict performance for the Black group.

Vey et al. (2001) conducted a meta-analysis to examine the predictive validity of the SAT for different race groups. The meta-analysis looked at educational, rather than occupational performance, and was based on large samples of over a quarter of a million students. They found operational validities for first-year GPA were between 0.3 and 0.4 for both the verbal and mathematics scores of the SAT for Asian, Black, Hispanic, and White American college-students. The only substantial difference found was a somewhat higher validity for the SAT in mathematics for the Asian group.

Overall, these large US studies and meta-analyses find very similar validities for all groups. However, there is a trend suggesting that validity may be a little lower for some minority groups.

Outside the USA, studies of differential validity are rare. A series of concurrent validity studies from SHL South Africa showed similar validities for the White minority ($r = 0.22$–0.31) vs. other ethnic groups ($r = 0.24$–0.29) for eight different cognitive ability tests including verbal, numerical, and non-verbal tasks. The sample was based on a total of 267 incumbents in moderate- to high-level jobs, of which 169 were White and 98 were Asian, African, or Coloured. However, one verbal test was valid only for the White group ($r = -0.01$ and 0.25 for the combined non-White ($N = 96$) and White ($N = 162$) groups, respectively).

Baron and Gafni (1989) compared the predictive validity of a cognitive ability battery, similar to the SAT. The criterion consisted of first-year university GPA scores for Jewish and Arab applicants to five different faculties in two Israeli universities. There were 2,185 Hebrew examinees and 496 Arabic examinees in total. The tests predicted performance equally well for both groups, apart from in one faculty where a slope difference in the regression equation for the two groups was found. However, there was consistent overprediction of GPA for the Arab group from the common regression line, despite their lower average score on tests (d ranged from 0.67 to 1.5 SD in the different faculties).

The general null effects could be due to sampling error, since differential validity studies do require large samples to have reasonable statistical power.

The synthetic differential prediction analysis technique, suggested by Johnson, Carter, Davison, and Oliver (2001), uses the common performance elements across job families to increase statistical power. Another technique, the selection validation index (Bartram, 1997) also seems to produce more accurate estimates of validity for small samples. The application of such approaches to differential data could be effective in extending our knowledge base in this area. Currently only the most available groups seem to have been studied.

Summary

The literature is consistent on the validity of cognitive ability tests for predicting job performance, training performance, and educational achievements. In the main, differential studies have found that this holds for all subgroups. However, trends for lower minority group validities are common. Findings from around the world generally support this. There are individual results suggesting differences in validity for particular groups and types of test. These may be due to sampling error, or may reflect a real trend. Where intercept differences are investigated the trend is to find that use of common regression lines overpredict (i.e., favour) lower scoring groups if there is a difference. However, meta-analyses focusing on correlations rather than full regression equations can mask these effects.

ADVERSE IMPACT

The actual adverse impact that will result from using test scores in a selection procedure is dependent not just on typical group differences in test scores, but the selection rule that is applied to the scores. The bigger the score differences and the lower the selection ratio, the more adverse impact will result. Sackett and Wilk (1994) and Bartram (1995) show the selection ratio for minority groups under a variety of different selection ratios and *d* conditions. It is clear that the impact of a test with even moderate score differences can be high when a very selective rule is used. The appropriateness or fairness of using test scores under these conditions is an ethical judgment. In a growing number of countries there are legal constraints. The law in the USA and UK essentially requires a justification for using the test scores, which is typically interpreted as evidence from validation and/or job analysis. Stronger evidence is required to support greater impact. As a consequence, approaches to setting selection rules have been developed to reduce adverse impact.

An effective way to reduce adverse impact resulting from test score differences is to use within-group selection. The selection ratio is applied independently to each group. This is equivalent to within-group norming of test

scores and results in zero levels of adverse impact with minimal reduction in utility (Hunter, Schmidt, & Rauschenberger, 1977). This practice is not, however, 'group-blind', and for this reason is not always seen as fair. Few legislative frameworks that address fairness in selection would allow it, although recent South African legislation specifically includes this practice as an acceptable approach to reducing imbalances in employment.

An alternative approach is to lower the cut-off score used in selection. However, this will result in a much greater reduction in utility while generally not entirely removing the adverse impact (Sackett & Ellingson, 1997). It also requires the selector to find an alternative predictor for the large number of people who pass a lower cut-off score. This could be a benefit to the overall validity of the procedure but will involve greater investment in the selection process.

Test score banding has also been suggested as a method for reducing adverse impact. With this approach, fixed or sliding score bands based on the standard error of measurement are established around the test score, and anyone scoring within a band is treated as 'equivalent'. Other methods are needed for selecting individuals from within the test score band (Aquinas, Cortina, & Goldberg, 1998; Cascio, Outtz, Zedeck, & Goldstein, 1992; Cascio, Goldstein, Outtz, & Zedeck, 1996). The purpose is obviously to balance criterion-related validity with diversity. However, test score bands may define statistical differences in test scores that may not be actual differences in predicted performance. Overall, the impact is to allow lower scorers to pass the cut-off, and this will usually result in some diminution of adverse impact resulting from test score differences.

A fourth way to reduce adverse impact is to use alternative predictors that result in smaller group differences. A number of alternative predictors to cognitive ability have been identified that meet this criterion. These include interviews, biodata measures, physical ability tests, situational judgment tests, role plays, work samples, and some personality traits; these all show lower or even nonexistent Black–White differences (Bobko, Roth & Potosky, 1999; Campbell, 1996; Hough et al., 2001).

Huffcutt and Roth (1998) found that interview ratings for Blacks and Hispanics were on average only about one-quarter of a standard deviation lower than those for White applicants. Thus, interviews do not appear to affect minorities nearly as much as mental ability tests, and group differences for the interview appear to be much closer to actual differences in job performance than group differences for ability tests. They also found that highly structured interviews have smaller group differences than less structured interviews.

Personality measures have also shown smaller group differences than cognitive measures. Hough et al. (2001) reviewed a number of studies and found aggregate *d* values between 0 and 0.31 for Black–White comparisons on the Big Five dimensions and 0 and 0.11 for Hispanic–White comparisons. Baron

and Miles (2002) and Ones and Anderson (2002) found similar small differences in UK data comparing Black, Asian, and White samples.

Although many of these approaches have been shown to have good validity, the domains measured by these alternate predictors are typically related more to interpersonal skills, work style, leadership, supervisory skills, and conflict resolution than to cognitive ability (Campbell, 1996).

Sackett and Wilk (1994) calculated that an equally weighted composite of two uncorrelated scores, one with a 1 SD difference and the other with none, would have a 0.71 SD difference. Sackett and Ellingson (1997) model the reduction in adverse impact that might accrue when a predictor that shows a large group difference is combined with one with a lesser *d* value to form a composite measure. They show that the magnitude of the effect will depend both on the *d* value of the two initial variables and the correlation between the variables. Pulakos and Schmitt (1996) found that, in an employment context, a composite of a verbal ability test, a situational judgment test, a structured interview, and a biographical data measure produced a difference of 0.63 SD, whereas a composite of the three tests without the verbal ability test produced a smaller difference of 0.23 SD. Similar results were shown by Baron and Miles (2002) in a simulation of the combination of a personality instrument with cognitive ability tests based on a UK general population sample of personality questionnaire responses (OPQ32, SHL, 1999) and a real selection rule used by an employer; adverse impact was reduced to well within the four-fifths rule even when only 30% of the sample was selected.

However, Ryan, Ployhart, and Friedel (1998) warn that the actual reduction in adverse impact will depend on the distribution of scores in the different groups as well as the correlation between the added and original predictors. Deviations from a normal distribution near the cut-off point found in their samples of 4,172 firefighter and police-officer applicants resulted in a much smaller reduction in adverse impact than expected from simulations of the composite approach.

Schmitt, Rogers, Chan, Sheppard, and Jennings (1997), in a study using the Monte Carlo approach, found that the validity of a composite of alternate predictors and cognitive ability may exceed the validity of cognitive ability alone, as well as reducing the size of subgroup differences. However, Sackett, Schmitt, Ellingson, and Kabin (2001) warn of the possibility of increasing group differences when the composite is a more reliable measure of an underlying characteristic for which group differences exist. If the additional predictors are relevant for the job, sufficiently different, and have small differences, composite selection methods offer the prospect of increased validity as well as smaller group differences. Bobko et al. (1999) estimated, in their matrix of relationships between cognitive ability measures, alternative predictors, and job performance, that composites of cognitive ability, biodata, interviews, and conscientiousness would produce a validity of 0.43 with 0.76 SD difference. The validity estimate is 0.13 higher and the group

difference figure is 0.24 lower than their estimate for cognitive ability alone.

Assessment centres could be seen as operating in a similar way by combining multiple predictors. Goldstein, Yusko, Braverman, Smith, and Chung (1998) found a Black–White difference of 0.4 SD for a composite score across all exercises. Hoffman and Thornton (1997) found an ability test alone had only slightly higher validity than a full assessment centre procedure, but the assessment centre had minimal adverse impact whereas the ability test showed typical group differences. However, in many cases, assessment centres reflect dimensions outside the cognitive domain. Therefore the reductions in group differences may be due to the lack of relationship between the alternative predictors and the main constructs tapped by traditional ability tests.

Summary

Supplementing cognitive ability tests with additional measures of other skills that typically show smaller group differences offers an alternative to just lowering cut-off scores in an effort to reduce adverse impact. Composites can both increase validity and temper adverse impact. However, in a number of studies, this approach has been less effective than anticipated suggesting that it is not a panacea, and the validity of alternative measures may not generalize as extensively as cognitive ability.

EXPLANATIONS FOR GROUP DIFFERENCES

Despite the consistent and large effects of group differences on cognitive ability tests, very little headway has been made in explaining or reducing the variance. Neisser et al. (1996) reported on the conclusions of a task force of the APA which looked at different issues surrounding intelligence. They review the evidence for social and biological causes of difference but do not find any body of evidence conclusive. We will consider a number of possible causes: those relating to the test-taker including background and education; emotional factors such as candidate motivation and anxiety; differential approaches to completion of tests, and factors relating to test design and administration.

Test-taker Background and Experience

One potential explanation of score differences could be the difference in experience of people from different groups, particularly while growing up. Both in the USA and the UK, members of ethnic minority groups are more likely to be among those of lower socio-economic status than those of the

White group. This pattern repeats itself around the world, with lower performing groups, even when not the minority, disproportionately belonging to less privileged strata of society. This of itself might reduce opportunities to develop due to poorer educational opportunities, lack of high-performing role models, or economic deprivation. For example, Schmitt, Sacco, Ramey, Ramey, and Chan (1999) found that parental employment was associated with positive changes in social and academic progress.

There are enough examples of individuals and groups performing well despite unpromising circumstances to suggest that this cannot be the only explanation of differences. However, test-taker background and experience is such a pervasive factor that it is highly likely to have some impact.

Socio-economic status (SES)

US data. Historically, Black Americans have had less wealth, more menial jobs, and less access to education than White Americans on average. In 1989, 27% of Hispanics were under the poverty level, in contrast with 12% of non-Hispanics (Rodriguez, 1992). Black and Hispanic students are more likely to come from families with lower parental education and less income (Camara and Schmidt, 1999). Adelman (1999) found a correlation of 0.37 between SES and a composite measure of academic achievements, including the SAT; a similar correlation of 0.32 was found between IQ and parental SES by White (1982). Schmitt et al. (1999) found that parental income and education were related to various school outcomes. Thus, considering the inequities minorities have suffered through poverty, discrimination, years of tracking into dead-end educational programmes, lack of access to advanced courses, poor facilities, overcrowding, poorly qualified teachers, and low expectations (Camara and Schmidt, 1999; Kober, 2001), it is reasonable to hypothesize that some of the differences seen in the test performance of Hispanics and Blacks may be related to low SES.

However, Camara and Schmidt (1999) show that, in addition to a general effect of SES on SAT scores, within any SES or parental education band, Black and Hispanic students scored lower on the SAT than Asian and White students. On average, students of these minorities coming from families with the highest levels of income and parental education still lag behind White and Asian students from families with moderate income and education. Similar patterns are found on non-test measures such as school grades and class rank. Findings such as these suggest that it is not just lower SES that is affecting the decreased scores of ethnic minorities on cognitive ability tests.

However, there are limitations to the measures of SES used in most studies. They are often very broad categorizations that band substantially different levels together and fail to capture factors such as large gaps in accumulated wealth and financial assets that persist after controlling for

education and income (Oliver & Shapiro, 1995). Research has found that White families often have three or four times more accumulated wealth and financial resources than minority families at the same income level (Belluck, 1999); hence more research is needed with more specific and detailed measures of SES in order to assess the strength of correlation with test performance.

UK data. The *Educational Inequality* report by Gillborn and Mirza (2000) examines differences in attainment at school by race and social class. Two social class categories are identified from parents' occupation: manual and non-manual background, where the former is taken as roughly equivalent to working class and the latter to middle class. The results of the research show that generally pupils from non-manual backgrounds have significantly higher attainments than their peers from manual households. Differences in scores between different social class groups are around twice the size of overall race differences in achievement. Population census data show that members of ethnic minority groups are more likely to have lower SES.

However, as with the US data, trends within social class and ethnic groups show that SES does not account for all of the observed race differences in performance. For African Caribbean pupils the social class difference is much less pronounced, with children of non-manual backgrounds performing little better than those from the manual background. On the other hand, those of Indian origin and manual background perform better than expected.

It is clear from this research that social class factors are related to attainment within each ethnic group. However, as in the USA, social class factors do not override the influence of ethnic difference, and, while there are clearly class differences in educational attainment, social class does not account for the entire difference found between ethnic groups.

Education

Many of the socio-economic findings discussed above relate to educational achievement measures rather than cognitive ability test scores. There is a strong relationship between measures of cognitive ability and educational achievement. *g* predicts academic achievements better than anything else. Kaufman and Wang (1992) found a strong correlation between educational attainment and intelligence for Whites, Blacks, and Hispanics in the USA.

However, early deficits in educational achievements can lead to later deficits in educational opportunities. A child who does not learn to read during the first few years of school may never gain access to an academic study track. A school drop-out is unlikely to go on to further education. A lack of educational opportunities may cause stunted cognitive development, which could

account for the measured differences in performance on cognitive ability tests in employment or other adult contexts. In this section we consider patterns of difference in educational outcomes.

Roth and Bobko (2000) in the USA found that Black–White differences in GPA increased from junior to senior years. The average difference for seniors was 0.78 SD, favouring the White group. They also found that White means tended to rise as students progressed, whereas Black means were more stable. Kober (2001) discusses what is generally known in the USA as 'the achievement gap'. This is a consistent trend for minority group achievements to lag behind White achievements by an average of around two grade levels. She points to a variety of economic, school, community, and home factors to explain the gap and suggests that the fact that the size of the gap differs substantially in different states, and that it shrank noticeably during the 1970s and 1980s when concerted educational programmes were used to address it, means that it is not immutable.

Research in the UK also shows a relationship between race and educational achievement. Gillborn and Mirza (2000) find that standardized differences in achievement between groups increase as educational level increases. In all local education authorities that recorded sufficient ethnic data, Black pupils' position in school relative to their White peers worsened between the start and end of their compulsory schooling. At the start Black pupils were the highest attaining of the main ethnic groups, recording a level of success 20 percentage points above the average. This is in contrast to US findings where the achievement gap is present before children start school (Kober, 2001). However, in their GCSE examinations at age 16, Black British pupils attained 21 points below the average.

One theory that has been offered to account for this situation in the UK is that Black pupils are more likely to become alienated from school. Qualitative research has consistently highlighted ways in which Black pupils are stereotyped and face additional barriers to academic success. They are often treated more harshly in disciplinary matters and teachers have lower expectations of their Black pupils, assuming them to have lower motivation and ability. However, studies have also shown that despite these barriers Black pupils tend to display higher levels of motivation and commitment to education, and receive greater encouragement from their families to pursue further education (Gillborn & Mirza, 2000).

These differences are also seen in higher education. Dewberry (2001) found a correlation of 0.11 (equivalent to 0.22 SD difference) between minority group membership and degree class for UK law-trainees taking their bar exams. There were correlations of 0.13 and – 0.11, respectively, between minority group membership and attendance at a highly selective university ('Oxbridge') or at a college that had only recently received university status. Thus minority trainees had lower university achievement scores and were more likely to have attended a less prestigious institution.

Summary

Pervasive differences in educational achievements between groups are evident. While there is some evidence that they can be moderated or even eliminated by appropriate interventions, the educational deficit patterns of minority groups found throughout the educational process may persist into occupational settings and indeed account for some of the differences found later on.

Perceptions, Motivation, and Anxiety

A substantial body of work looking at the ethnic differences in individual attitudes to test-taking, such as motivation and anxiety levels, has built up over the last decade. This section will focus on how these factors mediate group differences in scores.

Ryan and Ployhart (2000), in their comprehensive review of candidate perceptions, identified several factors, not related to ability, that might influence test performance. We will review the main findings, focusing upon test motivation, test anxiety, and candidate perceptions of testing situations. The factors are listed below:

- test motivation;
- test anxiety;
- belief in tests;
- perceptions of job-relatedness of test;
- perceptions of predictive validity and face validity;
- perceptions of procedural justice;
- test ease;
- prior test experience.

Test motivation

One framework for understanding performance on cognitive ability tests suggests it is the product of two main factors: ability and motivation. Arvey, Strickland, Drauden, and Martin (1990) showed that individuals who complete tests for research purposes perform less well than those who have more to gain through performing well on the test. They compared the scores of applicants and incumbents and attributed the difference to test motivation. They concluded that racial differences in test scores might be related to test attitudes.

Arvey et al. (1990) developed a 60-item Test Attitude Survey (TAS) to examine attitudes toward testing, and the subsequent influence on performance. They used it with 263 applicants to a financial-worker position, and found White Americans reported levels of test motivation the equivalent of 0.26 SD higher than Black Americans. There was also a positive correlation

between levels of motivation and test performance ($r = 0.20$; equivalent to a d of 0.39 SD), suggesting that racial differences in test motivation may to some extent influence test results.

Chan, Schmitt, DeShon, Clause, and Delbridge (1997) studied the test perceptions of a sample of 210 undergraduates from a US university between two administrations of parallel forms of a cognitive ability battery and found larger effect sizes. Motivation was engendered by offering students scoring in the top 40% additional payment. Test-taking attitudes were measured using the TAS. They found Whites reported levels of motivation 0.45 SD higher than Black participants. Correlations of 0.37 and 0.40 were found between motivation levels and the first and second test administrations, respectively. They suggest that motivation could account for some of the 0.87 SD difference between the groups in test scores. Ryan, Ployhart, Greguras, and Schmit (1998) and Schmit and Ryan (1997) found similar differences in motivation between Whites and Blacks among applicants for firefighter and police-officer positions, respectively.

In contrast to these US findings, Mains-Smith and Abram (2000) studied 579 UK school-students, using an adapted version of the TAS. They found significantly higher levels of test-taking motivation (0.3 SD) among the 353 ethnic minorities (mainly of Asian origin) than the White group and a negative relationship between motivation and test scores. This opposite effect also accounts for some of the 0.5 SD test scores difference. The higher motivation finding for the ethnic minority group mirrors the educational findings discussed earlier (Gillborn & Mirza, 2000).

Test anxiety

Levels of test anxiety may have a moderating influence over test performance and these have been shown to differ by race. Samuda (1975) found that more Black American students than White American students suffered from debilitating levels of test anxiety (i.e., anxiety levels that are so high that they have a negative influence on test results). This result has been replicated by Clawson, Firment, and Trower (1981), Payne, Smith, and Payne (1983) and Rhine and Spaner (1983).

Arvey et al. (1990) found no difference in test anxiety between Black American and White American candidates but there were negative correlations between test anxiety and three tests of cognitive ability, ranging from -0.21 to -0.47. Schmit and Ryan (1997) also found a negative correlation between pre-test levels of anxiety and test performance, using a student sample ($r = -0.11$; $N = 323$). Ryan et al., (1998) found higher levels of test-taking anxiety among Black firefighter candidates than for the White majority.

Ryan (2001) suggests that higher levels of test anxiety could be related to more negative self-evaluations, more task-irrelevant thinking, decreased

attention to task-relevant cues, debilitating emotionality, and withdrawal of effort, which combine to produce lowered performance levels. Steele (1997) and Steele and Aronson (1995) suggest higher reported levels of test anxiety in Black candidates might be traced back to the 'stereotype threat' felt by some minority candidates. This is the fear of confirming a negative stereotype of Blacks through performing badly on the test. These studies have shown that stereotype threat occurs in experimental situations when group identity is made salient and can reduce the test scores of Blacks. Shih, Pittinsky, and Ambady (1999) found improved performance on a maths test for Asian women in conditions of stereotype threat linked to their ethnic origin, but lower scores for stereotype threat linked to their gender. However, Sackett et al. (2001) warn that the effect has failed to reproduce outside experimental samples.

Dion and Tower (1988) compared test anxiety for White and Asian American students at a Canadian university and found that Asians also typically show higher levels of anxiety, although they tend to outperform Whites in terms of test sores. They speculate that this may be the result of facilitating levels of stress and anxiety engendered by family pressure to succeed.

Mains-Smith and Abram (2000) also looked at anxiety levels in their study of British students. In this case their findings were similar to US studies. The pooled ethnic minority group ($N = 366$) reported higher levels of test anxiety (0.49 SD) than the White group ($N = 229$), and there was a negative correlation between levels of test anxiety and test performance ($r = -0.41$), showing that those who are more anxious do significantly less well on the test. Kurz, Lodh, and Bartram (1993) also report greater test anxiety among ethnic minority UK school-students taking tests as part of a research project.

A curvilinear relationship between anxiety and performance has been suggested (Anastasi & Urbina, 1997), and this may be one explanation of why anxiety studies produce less consistent results.

Test perceptions

In addition to motivation and anxiety, candidates' attitudes toward tests could mediate performance. This might include their belief in tests as effective measures of ability, perceptions of job-relatedness and relevance, perceptions of the appropriateness of including tests in a selection procedure, and the effectiveness of doing so (i.e., test validity). Group differences in attitudes could therefore contribute to score differences.

Chan et al. (1997) asked students to rate the face validity of a series of cognitive tests they had completed for a managerial job based on a list of typical skills required for the role. Black students rated the tests less face-valid than White students by 0.28 SD. Ratings of face validity showed a correlation of 0.31 with test performance. Structural equation modelling suggested that perceptions of face validity impacted performance indirectly

through its effect on test motivation. Ryan, Sacco, McFarlane, and Kriska (2000) reported that Blacks applying for positions as police-officers had more negative perceptions of testing processes generally than did their White majority counterparts.

Chan (1997) reported similar findings showing that student ratings ($N = 241$) of predictive validity of a battery of cognitive tests correlated with both race ($r = 0.18$) and scores on a cognitive ability test ($r = 0.14$).

Chan and Schmitt (1997) showed that the difference between Black and White college-students' perceptions of face validity was smaller for a video-based presentation of a situational judgment test relative to a paper-and-pencil version. However, for both versions Whites rated face validity as higher.

Hough et al. (2001) review attempts to relate these negative perceptions of the testing process to the higher drop-out rate of Blacks in selection processes but conclude that there is no evidence of any strong relationship. In contrast to US findings, Mains-Smith and Abram (2000) found their mixed group of British ethnic minority school-pupils had a greater belief in tests than the White group. This may help explain higher motivation levels for this group.

Summary

The US research shows consistent relationships between race, motivation, anxiety, and perceptions of tests, which suggests that these factors could account for some part of typical Black–White score differences on cognitive ability tests. There are far fewer studies looking at other ethnic groups, and those that there are do not follow the pattern seen in Black–White studies. Attitudes and feelings about testing processes could well differ from group to group and country to country. Further research is needed to try to understand how these factors interact and impact on test scores. Such studies might lead to effective interventions to reduce score differences a little.

Preparation and Coaching

It is generally considered good practice to make some provision for all candidates to arrive at the testing situation equally prepared to be tested. There is some belief that this may in some way reduce the differences between candidates from different groups and allow them to exhibit their true levels of ability (e.g., Anastasi & Urbina, 1997). This suggests that ethnic minority performance improves significantly more than White majority candidates' performance through preparation and coaching interventions.

Test orientation programmes are widely used, particularly in public sector selections in the USA (Hough et al., 2001), and the provision of practice materials is strongly advocated among experts in testing in the UK (e.g., Cook, Mains-Smith, & Learoyd, 2000; Toplis, Dulewicz, & Fletcher,

1997; Wood & Baron, 1993). The purpose of both these interventions is to provide information on the nature of the tests and, through this, increase belief in the tests and test motivation, and reduce test anxiety.

Coaching interventions are more intensive and therefore less frequently used. These often extend to multiple sessions and typically include more detailed input on test-taking skills. They may also include material intended to help develop the skills being measured by tests (e.g., quantitative reasoning). Clause, Delbridge, Schmitt, Chan, and Jennings (2001) found that test preparation activities (meta-cognition and learning strategies) were associated with higher test performance.

Much of the literature on practice and coaching comes from the educational domain. Sackett, Burris, and Ryan (1989) noted that educational studies generally find a positive effect on cognitive ability test scores of coaching programmes, but they find some evidence for smaller effects for lower ability attendees—or, in other words, those of higher ability may benefit more from preparation and coaching. Thus orientation programmes could increase rather than decrease group difference findings. Ryan, Schmidt, and Schmitt (1999) found minority scores for entry-level manufacturing jobs increased by 0.15 SD with an orientation programme. However, similar, or sometimes larger, differences were found for the majority White group. Powers (1993) summarized general findings relating to the SAT and concluded that there were greater effects for the quantitative scores than the verbal scores, and that lengthening a programme has a greater impact but the effect does asymptote. The review also emphasizes the importance of considering self-selection. Studies that do not take this into account often find effect sizes many times greater than those that do.

Johnson and Wallace (1989) looked for differential effects on individual item types in quantitative SAT items from a coaching programme aimed at Black students. They found only modest effects but some indication of more impact on items requiring some mathematical knowledge and a higher completion rate for candidates following coaching—suggesting that test-taking skills had been improved. The lack of a White comparison group means that it is unclear whether these are general coaching effects or are likely to have an impact on group differences.

Ryan et al. (1998) examined a coaching programme for applicants to firefighter positions in the USA offered by the hiring organization. They found higher participation rates for Black Americans than White Americans. There was no significant difference in test scores between those who attended and those who did not, nor were there any differences in motivation or anxiety when they were assessed immediately after the operational test. There was also no evidence of differential benefits to White or Black attendees.

In the UK, Fletcher and Wood (1993) report on a coaching intervention with a small number of applicants, mainly of Asian origin, for a position with a railway company. Results indicated a clear improvement in both test-taking

motivation and performance. However, the small size of the sample and the lack of a White comparison group limit conclusions from the study. Kurz et al. (1993) studied 163 British school-students offered various preparation opportunities before testing including a control group with none. For verbal and numerical tests there was no change in the score difference between Whites and ethnic minorities (mainly Asian) students. However, for a clerical speed and accuracy task the group difference reduced from over 1 SD in the control group to 0.38 SD for those offered preparation opportunities. Further analysis suggested that the score improvement following practice was in accuracy of responding more than speed. There was some indication that a few ethnic minority students in the control group had misunderstood the test instructions for the clerical task and they had substantially lowered the mean for the group. This is consistent with Fletcher and Wood's (1993) suggestion that their group of older Asians needed up to twice the time typically allowed to assimilate test instructions.

Summary

Overall there is little support for the use of preparation and coaching to reduce group differences. However, as usual the majority of the studies were carried out in the USA where children are accustomed to standardized testing from their school. A larger impact might be expected with groups who were unfamiliar with testing practices. This is more likely to be the case in other parts of the world, where testing is not so well embedded in the educational culture.

Test-taking Approach

There are many facets of test-taking strategy, but relatively little research in this area. The relative importance of speed and accuracy in responding may differ for different groups, and this may lead to more and less effective testing strategies that contribute to group differences with timed tests. Cultures differ in the way they value pace of work and risk-taking (e.g., Trompenaars & Hampden-Turner, 1993). Previous experience with tests may help candidates develop more sophisticated and effective test-taking approaches. We consider three elements in test-taking approach—speed, accuracy, and guessing.

Speed

In the West, speed of performance is highly valued and there are jobs (e.g., air-traffic-controllers) where speed of operation is essential and many others (e.g., programming) where pace impacts on output. However, it can be argued that unless speed of work is a key element in job performance,

restrictive time limits may bias test scores against slower performing candidates or those who take more time to check answers. A person with a cultural background that places less value on speed of performance could underperform on a test without more generous time limits (Sackett et al., 2001).

Schmitt and Dorans (1990) reported that Hispanic students tended to reach the end of the verbal sections of the SAT less frequently than White students with comparable total scores. Llabre (1991) studied Hispanic students and concluded that most research indicated that increasing time limits would differentially enhance their performance. Llabre and Froman (1987) showed that time spent on individual items correlated less with item difficulty for Hispanics than for non-Hispanics. This may have been due to difficulties with the English language, or to a lack of test sophistication and unfamiliarity with budgeting time in tackling items. There has also been some evidence showing an interaction between test speededness and test anxiety for Hispanic students which was not present for non-Hispanic students, (Rincon, 1979, in Pennock-Roman, 1992).

Dorans, Schmitt, and Bleistein (1992) looked at differential speededness on SAT tests and found that Black students had substantially lower completion rates for items at the end of the test. However, an increase in the amount of time allowed per item seems to increase group differences from 0.83 SD to 1.12 SD (Evans & Reilly, 1973). Wild, Durso, and Rubin (1982) found that increasing the time allotted may benefit all examinees, but did not produce differential score gains favouring minorities and often exacerbated the extent of group differences. This was the conclusion of Sackett et al. (2001) in their recent review. They suggest that increasing the amount of time allowed to complete a test often increases subgroup differences, sometimes substantially.

Kurz (2000) studied UK college samples and found very small differences in speed (measured by number of items attempted) and accuracy (proportion of items correct) between a White and a mixed ethnic minority group on relatively generously timed verbal tests, with Whites completing slightly more items slightly more accurately. However, no differences were found on highly speeded numerical tests. Although the ethnic minority group was quite large (148 out of a total sample of 930), it was made up largely of foreign students many of whom had only a moderate command of English. te Nijenhuis and van der Flier (1997) found their immigrant groups completed fewer items than the majority Dutch group. Pennock-Roman (1992) points out that ability is likely to be underestimated when examinees are tested in their weaker language.

Cook (1999a) found that the trend across the subtests of the computer-administered BARB was for the ethnic minority candidates ($N = 581$) to attempt fewer questions and take significantly longer to respond to each question when compared with White applicants across most of the six subtests. Asian applicants ($N = 154$) attempted fewer items and had longer

response times on all but two of the subtests—SIT subtest (a test of Semantic Identity, or 'odd one out') and Number Distance Task (a test of working memory and basic numeracy involving speed and accuracy). The response times of the Black group ($N = 410$) were similar to White applicants on only one task—the Rotated Symbol Task (a test of spatial orientation or mental rotation). In this case, language was not the issue as only 2.7% of the ethnic minority sample reported anything but English as their primary language. Mains-Smith and Abram (2000) examined the RT and found similar results. The White group attempted significantly more questions per subtest than the ethnic minority group in the time given. However, these patterns may reflect response patterns of lower-scorers generally, therefore further research with a matched control would be helpful.

Overton, Harms, Taylor, and Zickar (1997) found that candidates took longer to complete test items that were closer to their ability levels. Candidates with lower abilities will therefore tend to spend longer on questions earlier on in the test. This suggests that the slower response rate may be due to lower ability in the test-taker rather than vice versa.

Accuracy

There are even fewer findings relating to differential accuracy on complex items and, when they exist, they can be confounded by speed effects. In the USA Steele and Aronson (1995), for example, found a marginal tendency for Black participants to evidence less accuracy than Whites on a 30-min test composed of items from the verbal GRE. Cook (1999a) found that the trend across the subtests of the BARB was for ethnic minority candidates to answer questions incorrectly more frequently. Mains-Smith and Abram (2000) examined the Royal Navy RT and found that, despite attempting slightly fewer questions, the ethnic minority group still had proportionally more wrong answers per item attempted than the White group.

Research on specific measures of clerical speed and accuracy may be relevant here. For example, Schmitt et al. (1996) cite Department of Defence data collected in 1980 using results from a clerical speed and accuracy test within the Armed Services Vocational Aptitude Battery (ASVAB). These data showed differences favouring Whites of 0.95 SD from Blacks and of 0.65 SD from Hispanics. These results were from military samples and may therefore have restricted generalizability. A small general sample showed a much lower difference (0.15 favouring Whites: Schmitt et al., 1996). Hough et al. (2001) find a Black–White difference of 0.35 SD and 0.38 SD for the Hispanic–White comparison. The SHL Group's UK data show variability around half a standard deviation on clerical speed and accuracy tests. However, whether scores on clerical speed and accuracy tests are related to the speed and accuracy with which other tasks are performed needs to be determined. If these findings generalize to general

test-taking style, they are consistent with the accuracy findings for more complex items.

However, as with the speed findings, it is not clear whether these differences are a cause of lower scores or are, perhaps, a typical response style for lower-scorers from any group. Score level and speed and accuracy are confounded in many of these studies.

Guessing

Guessing behaviour could differ for different groups. Cultural attitudes might affect the value a candidate places on answering accurately. A high value might encourage someone to check answers more thoroughly during a test and deter guessing when uncertain. Both of these effects would tend to reduce scores on standardized multiple-choice tests with standard scoring. Use of correction for guessing might help to reduce any differences found in this way, but only controls for blind-guessing. Candidates who guess wisely (e.g., when one or more answer options can be ruled out) can still benefit from guessing, even when a correction is applied. No studies of this area were found.

Freedle and Kostin (1997) found that Blacks tended to omit more questions than Whites, which suggests a lower propensity to guessing. In contrast, Dorans et al. (1992) suggest fewer omitted answers for Hispanic respondents compared with Whites. Both these studies looked at SAT results. Further investigations of these trends in the occupational sphere would be useful.

Jaradat and Sawaged (1986) suggested a 'subset selection technique' in which candidates are instructed to mark all the answers they think might be correct. Where they know the answer, a single response can be marked. Where they are unsure, but can rule out one or two options, only the remaining ones are marked. They suggest that the approach does not favour high-risk takers and if anything enhances the reliability and validity of the test. Interestingly this work was carried out in Jordan, a very different cultural environment from that used in most Western studies. Further research into the modification of the response process could lead to reductions in score differences.

Summary

Research has been limited and inconclusive in the area of differential test-taking approaches. It appears from this limited research that ethnic minority groups may complete cognitive ability tests marginally slower and less accurately than White groups, although this is by no means undisputed. Attempts to increase the time allotted for the test in order to reduce the pressure on ethnic minority candidates have often benefited all candidates. The number

of items completed, and the number of items completed accurately, may increase for all groups. Thus, the increased time available to test-takers often exacerbates the extent of group differences.

Some contradictory findings may relate to the differential speededness of the tests studied and ceiling effects in scores before the experimental manipulation.

Test Design

Many critics suggest that differences in performance between groups is a result of tests that reflect the reasoning processes, definitions of intelligence, cultural assumptions, linguistic patterns, and other features of the test-writers. In the USA, and to a large extent around the world, this is the dominant White culture, influenced as it is by rationalism, Western European cultures, and Judeo-Christian thought. In this section we review some of the mechanisms suggested and studies that attempt to check whether these do affect group differences. Helms (1992) is perhaps one of the more coherent critics and, while researchers have begun to address some of the hypotheses she raises, there are still only a few relevant studies. Many of these are performed in a research context and investigate trends in small samples of college students. Larger studies with more realistic occupational groups would be desirable.

Standardization on White groups

Harrington (1988) pointed out that typically tests are standardized using predominantly White samples. It could be that this process is tending to select items and create tests that favour the White group. One study that examines this is by Hickman and Reynolds (1986), who tested this hypothesis by creating two forms of a cognitive battery for children standardized on majority Black and majority White samples, respectively. They found no difference in score patterns for the two forms. Similar null results were found by Fan, Willson, and Kapes (1996). Jones and Raju (2000) do manage to find some effects with an Item Response Theory (IRT)-based approach—their results are discussed in the next section.

Differential item functioning

If there are cultural factors that make tests and items differentially difficult for some groups, it is likely that these load more on some items than on others. Attempts to identify inappropriate items through reviews were not always successful. A Wechsler Intelligence Scale for Children (WISC) item identified as unfair to Black children turned out to be relatively easier for them. ‘Culture fair’ tests, such as the Cattell Culture Fair Intelligence Test

(Cattell & Cattell, 1960–1961), also failed to reduce group differences. In the 1980s effective statistical methods for identifying individual items in tests that might be more difficult for a particular group were developed (Dorans & Kulick, 1986; Holland & Wainer, 1993). This is generally referred to as differential item functioning or DIF.

DIF findings are often difficult to interpret, because of the many comparisons required for a single test. Type-1 errors can be a major problem unless significance levels are increased to such an extent that only the very largest effect sizes can be identified. Generally studies have found a small number of DIF items in tests of cognitive ability. These items do not always favour the higher scoring group, and removing them has only a small impact on overall group differences. However, the use of DIF techniques, together with focused reviewing for fairness, has become a standard part of test development processes. While this may have had only a minor impact on group differences, it has certainly led to more acceptable content in modern tests compared with those developed in the past.

Positive DIF findings are often difficult to explain, but have sometimes been related to familiarity with item content and the verbal complexity of the items. However, beyond the influence of having the test language as one's primary language, there is little information about how cultural differences affect test performance (Sackett et al., 2001). Scheunemann and Gerritz (1990) studied verbal items from the SAT and GRE. They found mixed effects across the two tests, but there was a trend for Blacks to find items with science-based content more difficult than Whites. Freedle and Kostin (1997) showed that Black examinees were more likely to answer difficult verbal items on the GRE and the SAT correctly when compared with equally able White examinees, but the Black examinees were less likely to get the easy items right. Freedle and Kostin (1997) suggested that this might be because the easier items possessed multiple meanings more familiar to White examinees, whose culture was more dominant in the test items and the educational system. The use of homographs (words that have more than one meaning for the same spelling) in tests has also been cited as a feature, although when non-native English-speakers were removed from the analyses, few DIF items remained, suggesting that the differences were due to language problems (Schmitt & Dorans, 1990).

Mains-Smith and Abram (2000) looked at DIF on the UK Navy tests but found no consistent explanation for items flagged. Removing flagged items had a negligible impact on test score differences.

Recently Raju and colleagues have developed a new technique (DFIT) for comparing differential functioning both at the item and the test level (Jones & Raju, 2000; Oshma, Raju, & Flowers, 1997; Raju, van der Linden, & Fleer, 1995). This combined IRT-based approach identifies differences in scores at different ability levels when the test is standardized using the majority and minority data, having first removed potential DIF items. As the procedure

looks at differences by score it is possible to focus on those differences around the cut-off score being used in a selection process, rather than average differences. Thus items that impact differences most in this score range can be removed.

Construct equivalence

One interpretation of positive DIF findings would be that those items were measuring different constructs for the groups compared. If enough items are implicated, the test itself might be seen as measuring a different construct. This is part of the Cleary definition of fairness; that is, tests should be measuring the same thing for every group (Cleary, 1966). Consideration of equivalence of constructs is part of the work of a test-developer and is rarely published in the peer-reviewed literature. Wing (1980) reports similar reliabilities for all groups for a battery of cognitive ability tests.

Schmitt and Mills (2001) examined the intercorrelations of two series of measures for majority and minority job-applicants. Structural equation modelling showed that while the structure of the measures from a simulation were similar for the two groups, scores from a set of more traditional paper-and-pencil tests showed greater variance for the Black group compared with the White group. Hattrup, Schmitt, and Landis (1992) looked at the factor structure of a series of six tests among applicants for entry-level manufacturing posts for different subgroups. Structural equation modelling revealed that the same models showed best fit in all subgroups, but in general fit was better for White groups than for Black or Hispanic applicants. te Nijenhuis and van der Flier (1997) found similar structures for the Dutch GATB tests for the majority and minority groups.

UK data from test publishers suggests similar test reliabilities for ethnic minority and White groups. However, where differences do occur they tend to indicate lower reliability for ethnic minority groups. This is sometimes, but not always, related to lower score variance for these groups (SHL Group, 2002).

There is no strong evidence that there are differences in construct validity for tests for different ethnic groups. However, we found no systematic studies of equivalence in this area.

Cultural equivalence

Helms (1992) argues that cognitive ability tests lack cultural equivalence. She suggests they assess White *g* rather than African or Hispanic *g*, and therefore that Whites may be expressing their abilities in the biological or environmental styles of their group, whereas Blacks are not. Helms (1992) suggests that score differences may be due to a cultural bias inherent in the tests and lists a large number of hypotheses relating to ways in which

African-influenced cultural assumptions held by Black Americans might lead to poorer performance on standardized tests. In such an instance the test performance of Black test-takers would be more indicative of level of acculturation or assimilation to White culture than of level of cognitive ability.

Geisinger (1992) argues that from the Hispanic perspective cognitive ability tests may be culturally biased toward Whites. He suggests that Hispanics do not have a cultural knowledge of the mechanics of testing nor do they adhere to the belief that the testing enterprise provides the standard to assess performance. This puts a traditionally minded Hispanic at a disadvantage in not understanding the implications of tests for future life chances. Degrees of acculturation also vary across individuals—in becoming acculturated, a Hispanic learns and accepts the norms sanctioning test results as the standards by which rewards and opportunities are given.

The cultural-distance approach, as outlined by Grubb and Ollendick (1986), suggests that a subculture's distance from the major culture on which questions of a test are based and validated will determine that subculture's subscore pattern. Humphreys (1992) suggests that IQ test items measure such components as information, knowledge, and understanding, and that these components are culturally loaded in favour of White groups. Schiele (1991) argues that the African American epistemology is characterized by the spiritual, the rhythmic, and the affective dimensions of life, and that IQ tests are culturally biased in their focus on left-brain functions (analytical thinking) while ignoring right-brain functions (holistic and artistic thinking). Schiele (1991) also suggests that IQ tests should assess musical IQ, bodily IQ, and personal IQ in order to eliminate bias. Grubb and Ollendick (1986) found that although Blacks and Whites performed similarly on learning tasks, they performed differently on standardized IQ tests, possibly because of the loading of cultural influences on the latter measures.

Identifying cultural factors that influence test responses has been difficult, and much of the existing research suggests that cultural differences do not account for racial differences on cognitive ability tests (Jensen, 1998). It could be possible that this is because most examinations of DIF are *post hoc* although a priori hypothesis-testing of DIF has also not supported the cultural theory (Hough et al., 2001). Helms (1992) suggests that this is because no substantive theory of culture is being tested, and that existing studies of cultural equivalence assess Black acculturation not Black intelligence. To reach a definitive conclusion, research needs to be theoretical and hypothesis-testing, with a more specific and measurable definition of culture.

Social context

One example of how cultural values may influence the test performance of ethnic groups is in the social content of the test items. DeShon, Smith, Chan, and Schmitt (1998) suggest that abstract measures of ability tend to yield

some of the largest performance differences between Black and White Americans. Helms (1992) argued that cognitive ability tests such as these fail to adequately assess African American intelligence because they do not account for the emphasis placed on social relations and the effect of social context on reasoning in the African American culture.

DeShon et al. (1998) found that both racial subgroups benefited from reasoning items presented in a social context, but, contrary to Helms' (1992) hypothesis, the subgroup difference did not decrease and even increased slightly. Castro (2000) used the Everyday Problem Solving Inventory (EPSI), which includes items in social and non-social situations, to compare the scores of White and Black Americans. The sample was small, 97 in total, of which 54 were White and 43 were Black. The findings suggested that group differences were present in EPSI scores in social domains but not in non-social practical domains. DeShon et al. (1998) conclude that the absence of social context in paper-and-pencil cognitive ability tests is not responsible for the observed performance differences between Black and White Americans. They argue that this may be because, contrary to Helms (1992), recent research suggests that Black and White Americans do not differ greatly on perspectives of human nature and social relations.

Method of test presentation: alternatives to written tests

It has been suggested that assessments that are more interactive, behavioural, and aurally–orally-oriented tend to exhibit smaller score differences than paper-and-pencil cognitive ability tests. Sackett (1998) suggests that, as oral exercises have generally shown smaller ethnic group differences, using different media such as video or multimedia to present the test items could help reduce differences. There are a number of studies relevant to this hypothesis but most have substantial weaknesses, either in the equivalence of the exercises presented in different media, or in the validation evidence available for the new medium. Often, on examination, the alternative modality assessments hardly relate to cognitive ability. For example, cognitive ability tests are compared with situational judgment tests. It is impossible to know if any resulting differences between subgroups were due to the difference in test content, constructs measured, or the medium of presentation.

Chan and Schmitt (1997) attempted to separate out test content from test method. They produced video-based and equivalent paper-and-pencil forms of a situational judgment test. Performance was significantly higher on the video form and the Black–White score difference was substantially smaller with this method. Performance and reading comprehension ability were nearly uncorrelated in the video administration, whereas they were positively correlated with the paper-and-pencil method, indicating that reading comprehension accounts for a substantial portion of the race/method interaction effects on test performance. There were no validity results for this test.

Pulakos and Schmitt (1996) compared measures of verbal ability, using a video-based writing task and a traditional multiple-choice measure. They also found a reduction in group differences: Black–White differences dropped from 1.03 SD to 0.45 SD with similar findings for Hispanics. However, the video test was less valid ($r = 0.29$) than the traditional verbal test ($r = 0.39$).

Richmann-Hirsch, Olson-Buchanan, and Drasgow (2000) suggest that the use of multimedia assessments result in more positive reactions from test-takers. Chan and Schmitt (1997) also found that students rated the video-based method significantly higher in face validity and that Black students saw the video-based tests as much more relevant than the equivalent paper-and-pencil test.

In contrast to the above studies, where the measures studied relate more to context factors than cognitive ability, Sackett (1998) summarized research on alternatives to the Multistate Bar Examination (MBE), which is a multiple-choice test of legal knowledge and reasoning. The Black–White difference on the MBE is 0.89 SD. A written research test alternative, did not decrease this difference nor did a video-based alternative which was also created. This contained vignettes of lawyers taking action in different settings and required candidates to respond to factual questions with a time limit of 90 minutes.

Klein and Bolus (1982, in Sackett, 1998) examined a combination of job simulations that was used as an alternative to the traditional legal bar examination. The simulations took place over 2 days and consisted of 11 exercises; for example, delivering an opening argument or conducting a cross-examination. An overall Black–White difference of 0.76 SD ($N = 485$) was found, higher than would be expected for a typical simulation exercise. The difference on oral tests was smaller than for written tests (0.46 SD, and 0.84 SD, respectively). Unfortunately there is no information on correlations between the traditional examination and the assessment centre.

Dewberry (2001) studied law exams in the UK. He found that Whites outperformed Black and Asian groups across a series of different examinations. A number of these were presented as role-play exercises (e.g., presenting a case) rather than paper-and-pencil tests. Unfortunately no standard deviations for scores are quoted in the article, therefore it is impossible to tell whether group differences were smaller for these exercises. What was noticeable was that whereas the Asian candidates outperformed Blacks on written tests the Blacks performed better on the role-play exercises.

Sackett et al. (2001) suggest that in total the research to date does not indicate that changing to a video or other format will reduce the group differences completely if the cognitive load is maintained. Failure to separate test content from test method confounds research results. Cognitive complexity is also confounded in these studies, with the high cognitive load simulations all at the high complexity end of the spectrum.

Job relevance

It could be hypothesized that higher job relevance would make the tasks seem more congruent and improve motivation and acceptability of measures, thereby reducing adverse impact. Ramist, Lewis, and McCamley-Jenkins (1993) found that Black and Hispanic candidates performed better on the subject matter-relevant SAT II achievement test than the more abstract questions of the verbal or quantitative reasoning papers of the SAT. Hattrup et al. (1992) compared results from generic cognitive measures and alternative paper-and-pencil measures designed to have higher job specificity. They found generally similar measurement properties and construct validity among the more specific and the general measures, but there seemed to be no reduction in adverse impact among the applicant groups studied. No criterion-related validity evidence was presented.

Job simulations seem to have less adverse impact than traditional tests (Chan & Schmitt, 1997) and participants tend to react more positively to simulations. For example, the use of work sample tests has been found to reduce the levels of adverse impact in a selection process (Robertson & Kandola, 1982). Pulakos, Schmitt, and Chan (1996) also found that role-play work samples resulted in much smaller mean score differences between Blacks and Whites than traditional paper-and-pencil tests (0.58 SD and 1.25 SD, respectively).

Schmitt et al. (1996) conducted a meta-analytic review of the literature and found differences of 0.38 SD between Blacks and Whites on job sample tests. This is encouraging because many of the tests in the group were written paper-and-pencil measures, which often display large subgroup differences. There were no differences between Hispanics and Whites on average. However, Pulakos et al. (1996) compared Hispanic–White mean score differences on different work sample tests and found that Hispanics on average still scored lower than Whites on all the measures (0.37 SD).

Many of the studies of job sample approaches do not address the issue of construct equivalence. It is quite likely that some of the job samples and simulations studied were not measuring the same abilities as the traditional cognitive ability tests with which they were compared. Schmitt and Mills (2001) addressed this issue in their study. They created a computer simulation of a call-centre job as an alternative to a more traditional paper-and-pencil battery. The job simulation consisted of a high-fidelity telephone task designed to replicate a day in the life of a service-representative. Candidates were required to receive and handle a number of ‘customer’ calls. Six different competencies were assessed by two assessors.

The results from nearly a thousand job applicants indicated that the traditional tests and the simulation exercises measured similar but not identical constructs. Structural equation modelling suggested separate but highly correlated factors for the ratings and traditional test scores. The factor

correlations were higher for Black respondents than for White respondents ($r = 0.69$ vs. $r = 0.57$). Similar to the findings of Chan and Schmitt (1997), *d* values for the simulation ratings (0.04–0.45 SD) were substantially smaller than for the test scores (0.37–0.73 SD). Using latent factor scores corrected for unreliability reduced the overall difference from 0.61 to 0.30 SD. Criterion data were collected for a small part of the sample and the traditional tests showed superior validity (0.46 as opposed to 0.36 corrected for restriction of range). Thus the simulation was a slightly less valid predictor of performance with less adverse impact than the traditional paper-and-pencil tests.

Alternative measures of cognitive ability

A number of authors have studied situational judgment tests (SJTs). This is another approach to creating a very job-relevant measure, although typically these tests have a strong behavioural element and are not pure cognitive measures. Weekley and Jones (1999) conducted a study using two different SJTs. Participants were also asked to complete traditional cognitive ability tests. Some 2,000 employees of five different retail organizations participated in the first study. All worked in store-level jobs such as checkout counter, stocking, and general-assistant positions. Of the sample, 89.5% were White, 6.2% Black, and 2.2% Hispanic. Although the White group outperformed the other two groups on both measures, there were slightly smaller group differences for the SJT relative to the cognitive ability tests (0.85 SD rather than 0.94 SD for Blacks and 0.23 SD relative to 0.52 SD for Hispanics).

The second study was based on data for around 1000 hotel employees with 'guest contact' roles. Of these 61.3% were White, 11.2% Black, 19.9% Hispanic, 6.5% Asians, and 1.1% were Native American. The study used cognitive ability tests and situational judgment tests but again found smaller group differences for the SJT. The reduction in difference in these studies is similar to those suggested by Motowidlo and Tippins (1993) in their earlier study, and by Pulakos et al. (1996), who found that an SJT resulted in much smaller Black–White differences than a written test of cognitive ability (1.25 SD vs. 0.35 SD, respectively).

The validity of such SJTs has been examined by Clevenger, Pereira, Wiechmann, Schmitt, and Harvey (2001). They indicate that situational judgment inventories (SJIs, similar to SJTs) are a valid predictor of job performance. Relative to alternate predictors, such as job knowledge, cognitive ability, job experience, and conscientiousness, SJIs had superior validity to most. Subgroup differences on the SJIs were also less than those for cognitive ability. Weekley and Jones (1999) demonstrate similar results,

also suggesting that SJTs were related to performance more strongly than cognitive ability, which is consistent with the job knowledge perspective that cognitive ability influences performance through its effect on situational judgment. The correlation found between cognitive ability and situational judgment was 0.42.

Content validity might suggest that SJTs often attempt to measure some element of social judgment. This could be seen as cognitive problem-solving in a social context. Helms (1992) suggests that social relevance is likely to reduce group differences on tests. However, these tests can also be characterized as a 'low fidelity' measure of social behaviour rather than cognitive problem-solving. In this case it would be more appropriate to compare findings on SJTs with personality measures or interpersonal exercises. SJTs show group differences that are at the high end of findings for these kinds of measure.

It is likely that different situational judgment measures will have differential loadings of cognitive, social, and behavioural factors. It would be useful to determine how the resulting group differences and criterion-related validity are related to cognitive loading. Video-based tests are often more strongly related to situational judgment measures than traditional cognitive tasks. Sackett et al. (2001) emphasize the importance of understanding the exact construct under investigation.

Reasons for the lower adverse impact of job-relevant simulations such as work samples and SJTs may be increased motivation on face-valid measures and lower reading requirements. The method of testing may also be of importance; that is, tests that minimize reading requirements and the use of written verbal material are likely to decrease the size of subgroup differences for low complexity jobs.

The conclusion drawn by Schmitt and Mills (2001) is that simulations are an alternative to traditional tests and, if they measure the same constructs, may help minimize adverse impact and increase positive reactions in participants and candidates.

Another alternative approach is to look at different aspects of cognitive functioning. Barrett, Carobine, and Doverspike (1999) found that short-term memory tests result in smaller standardized differences between Blacks and Whites than a reading comprehension test. Verive and McDaniel (1996) in their meta-analysis show that short-term memory tests can also have good validity for at least some jobs ($r = 0.41$). In the UK, the Army BARB test focuses on working memory capacity and results in slightly smaller score differences than the more traditional Navy tests. Further examination of the relative validities of these approaches, compared with traditional tests, as well as investigation of the scope of validity generalization would be warranted. Helms (1992) suggests these abstract approaches should be less appropriate for non-White groups whereas these results suggest the opposite.

Summary

There has been a considerable amount of research into the issue of whether group differences on test performance may be occurring due to bias in the test design itself. Several alternatives to the traditional test have also been suggested, with the aim of reducing group differences while maintaining predictive validity.

However, studies examining the various possible explanations for test bias have often found only small effects. For example, research suggests that item bias may account for little or none of the subgroup difference, and there is rarely a consistent pattern of items favouring one group over another. Studies of cultural bias have not been conclusive and seem to suggest that cultural differences only account for a small proportion of group differences on cognitive ability tests. However, there are issues with the study of cultural factors that need to be resolved in order to address the question specifically.

Despite some promising findings for a number of studies investigating alternative methods of testing, group differences have often not been reduced when care is taken to match the constructs measured. More research is needed that accurately separates test content from test method to assess whether alternatives to the traditional cognitive ability test can reduce the subgroup differences at different levels of complexity.

CONCLUSIONS

Our review started by looking at the evidence of race group differences in test scores. We were not surprised to find consistent evidence of group differences, both in the US studies and around the world. Group differences are not a US phenomenon, but it is mainly in countries with relevant legislation that data are being collected. Both the more recent US studies and the international findings emphasize the variability of results in contrast to the oft-quoted one standard deviation difference. The size of the difference seems to depend on a number of factors: the nature of the ability tested, the group tested, and self-selection or pre-selection within samples. Language can also be a factor for some groups. Researchers need to take this into account, and studies of innovations to reduce score differences need to include traditional measures as controls, rather than relying on comparisons of effect sizes with the assumption of one standard deviation difference. Further research needs to consider the nature of the groups studied in terms of race or ethnicity, prior selection, and, if possible, social and educational background. Studies also need to cover both true experimental designs and fieldwork in real selection situations.

There is still consistent evidence of validity for cognitive tests. There are few significant differences in studies of differential validity, but we found few

recent studies in this area in the occupational sphere. Most report comparisons of correlations, so do not address the potential for over- or under-prediction of performance for one group. Overall, studies suggest that there is validity for all groups, but it may be a little higher for White groups, and common regression lines may over-predict performance for lower scoring groups. Where language is an issue, validity may be substantially reduced.

The degree of adverse impact in selection decisions flowing from group differences has been more extensively studied in recent years. A number of authors have published tables of predicted impact giving various selection ratios and observed differences. There has also been some effort to look at ways of using scores that will reduce the adverse impact flowing from tests. The most effective involve combining cognitive ability tests with other types of measure that have less adverse impact. The resulting selection process can have better validity than cognitive ability alone with less adverse impact. However, it is clear that alternatives need to be well chosen, as results do show some variability in the effectiveness of this approach.

We reviewed a number of streams of evidence in attempting to understand the source of group differences. While there is no single explanation of differences, there seem to be a number of factors that may work together to create large differences. There is consistent evidence that lower scoring groups belong predominantly to lower socio-economic strata and have fewer educational opportunities. These effects do not by any means account for all the difference, but do seem to be a substantial contributing factor. Another factor seems to come from candidates' emotional responses to the testing situation. There is evidence of group differences in test-taking motivation and anxiety, and in belief in tests as an effective selection tool. All these factors have been shown to impact on test scores. The studies we identified in this area are almost exclusively US-based. Studies from elsewhere are few, but are particularly interesting since international comparisons allow the separation of underlying group differences in these factors and the impact of socially influenced values. Further research needs to focus on the circumstances in which these group differences arise, and the interaction between factors such as motivation and anxiety, as well as interventions that might control these effects.

One attempt to reduce differences is through preparation and coaching programmes. However, the evidence that they are effective in this respect is less than convincing. They may well have a function in influencing perceived fairness of the selection process, but it is not clear that there is any greater benefit for lower scoring groups. In some ways this is surprising, because there are indications that test-takers from different groups do take a different approach to tests in terms of factors such as speed, accuracy, and guessing. This is another area where further research would be beneficial.

Another approach to reducing group differences has been through looking at the design and construction of tests. A number of hypotheses have been raised about ways to measure cognitive ability that would show smaller differences. However, carefully controlled studies typically result in null effects. Approaches such as couching test items in a social context or using oral presentation seem to have little effect when the cognitive load in the tasks is similar. What these studies have identified is a number of valid constructs and methods of measurement that are related to, but different from, cognitive ability. These range from situational judgment to short-term memory. There are a number of these approaches that, when job-relevant, show promise in providing valid selection with less adverse impact. The generalizability of the validity of cognitive measures make them attractive selection tools, but the social impact of their use should not be ignored.

It is gratifying that some inroads are being made into understanding ethnic group differences. We have identified some areas here that seem to explain some part of score variance between groups, and further research into these areas may well help our understanding and our ability to control group differences. More international findings would allow inferences regarding the generalizability and even the causes of effects, as well as helping to address issues of test fairness wherever tests are used. However, there is no room for complacency. Group differences are still not well understood and much further work is needed.

ACKNOWLEDGEMENTS

This work was funded by SHL Group plc and the Chemical Biological Defence and Human Sciences Domain of the MOD's Corporate Research Programme.

REFERENCES

Adelman, C. (1999). *Answers in the Toolbox: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment*. Washington, DC: US Department of Education, Office of Educational Research and Improvement.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th edn). New York: Macmillan.

Aquinas, H., Cortina, J. M., & Goldberg, E. (1998). A new procedure for computing equivalence bands in personnel selection. *Human Performance*, **11**, 351–365.

Arvey, R. D., Strickland, W., Drauden, G., and Martin, C. (1990). Motivational components of test-taking. *Personnel Psychology*, **43**, 695–716.

Baron, H., & Chudleigh P. (1999). Assessment: Is bias in the eye of the beholder? Paper presented at the British Psychological Society Test User Conference.

Baron, H., & Gafni, N. (1989). The examination of criterion related and item bias for Jewish and Arab examinees in Israel. Paper presented at the AERA Annual Meeting in San Francisco.

Baron, H., & Miles, A. (2002). Ethnic differences on personality questionnaires. Paper presented at the British Psychological Society Occupational Psychology Conference, Blackpool.

Barrett, G. V., Carobine, R. G., & Doverspike, D. (1999). The reduction of adverse impact in an employment setting using a short-term memory test. *Journal of Business and Psychology*, **14**(2), 373–377.

Bartram, D. (1995). Predicting adverse impact in selection testing. International Journal of Selection and Assessment, 3(1), 52–62.

Bartram, D. (1997). The Selection Validity Index (SVI): Measuring predictive validity without correlations. Proceedings of the British Psychological Society Occupational Psychology Conference, Blackpool.

Belluck, P. (4 July 1999). Educators investigate racial gap in achievement. *New York Times*, B1, B6.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, **52**, 561–589.

Camara, W. J., & Schmidt, A. E. (1999). *Group Differences in Standardized Testing and Social Stratification* (College Board Report No. 99-5). New York: College Entrance Examination Board.

Campbell, J. P. (1996). Group differences and personnel decisions: Validity, fairness and affirmative action. *Journal of Vocational Behavior*, **49**, 122–158.

Cascio, W. F., Goldstein, I. L., Outtz, J., & Zedeck, S. (1996). Twenty issues and answers about sliding bands. *Human Performance*, **8**(3), 227–242.

Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1992). Statistical implications of six methods of test score use in personnel selection. *Human Performance*, **4**(4), 233–264.

Castro, W. (2000). The assessment of practical intelligence in a multicultural context. *Dissertation Abstracts International, A (Humanities and Social Sciences)*, **60**(10-A), May.

Cattell, R. B., & Cattell, A. K. S. (1960-1961). *Cattell Culture Fair Intelligence Test*. Champaign, IL: Institute for Personality and Ability Testing.

Chan, D. (1997). Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *Journal of Applied Psychology*, **82**(2), 311–320.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, **82**(1), 143–159.

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions and test-taking motivation. *Journal of Applied Psychology*, **82**(2), 300–310.

Chung-Yan, G. A., Hausdorf, P. A., & Cronshaw, S. F. (2000). Selection differences between cultural groups: Canadian data in a large urban transit authority. Poster presentation for Canadian Psychological Association, Ottawa, Canada.

Clause, C. S, Delbridge, K., Schmitt, N., Chan, D., & Jennings, D. (2001). Test preparation activities and employment test performance. *Human Performance*, **14**(2), 149–167.

Clawson, T. W., Firment, C. K., & Trower, T. L. (1981). Test anxiety—another origin for racial bias in standardised testing. *Measurement and Evaluation in Guidance*, **13**, 210–215.

Cleary, A. (1966). *Test Bias: Validity of the Scholastic Aptitude Test for Negro and White Students in Integrated Colleges* (Research Bulletin RB-66-31). Princeton, NJ: Educational Testing Service.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, **86**(3), 410–417.

Cook, R. D. (1999a). *Evaluation of Adverse Impact for the British Army Recruitment Battery Test (BARB)* (DERA/CHS/MID/CR980264/1.0). Farnborough, UK: Defence Evaluation and Research Agency.

Cook, R. D. (1999b). *Evaluation of Adverse Impact for the Royal Navy Recruiting Test (RT)* (DERA/CHS/MID/CR990090/1.0). Farnborough, UK: Defence Evaluation and Research Agency.

Cook, R. D., Mains-Smith, N., & Learoyd, K. (2000). *Best Practice in Military Selection (Other Ranks) Technical Report* (DERA/CHS/MID/TR000274/1.0). Farnborough, UK: Defence Evaluation and Research Agency.

Deakin, P. (2000). *Able Fairness and Equal Opportunities*. Oxford: Oxford Psychologists Press.

DeShon, R. P., Smith, M. R., Chan, D., & Schmitt, N. (1998). Can racial differences in cognitive test performance be reduced by presenting problems in a social context? *Journal of Applied Psychology*, **83**, 438–451.

Dewberry, C. (2001). Performance disparities between Whites and ethnic minorities: Real differences or assessment bias? *Journal of Occupational and Organizational Psychology*, **74**, 659–673.

Dion, K. L., & Tower, B. B. (1988). Ethnic differences in test anxiety. *Journal of Social Psychology*, **128**, 165–175.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance in the Scholastic Aptitude Test. *Journal of Educational Measurement*, **23**, 355–368.

Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, **29**, 309–319.

Evans, F. R., & Reilly, R. R. (1973). A study of test speededness as a potential source of bias in the quantitative score of the admission test for graduate study in business. *Research in Higher Education*, **1**, 173–183.

Fan, X., Willson, V. T., & Kapes, J. T. (1996). Ethnic group representation in test construction samples and test bias: The standardization fallacy revisited. *Educational and Psychological Measurement*, **56**, 365–381.

Fletcher, S., & Wood, R. (1993). The efficacy of coaching in test-taking behaviour: A preliminary investigation. Paper presented at the British Psychological Society Conference, Blackpool.

Freedle, R., & Kostin, I. (1997). Predicting Black and White differential item functioning in verbal analogy performance. *Intelligence*, **24**, 417–444.

Geisinger, K. F. (1992). *Psychological Testing of Hispanics*. Washington, DC: American Psychological Association.

Gillborn, D., & Mirza, H. S. (2000). *Educational Inequality. Mapping Race, Class and Gender: A Synthesis of Research Evidence*. London: Office for Standards in Education.

Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology*, **51**, 357–374.

Gottfredsen, L. S. (1988). Reconsidering fairness: A matter of social and ethical priorities. *Journal of Vocational Behavior*, **33**, 293–319.

Grubb, H. J., & Ollendick, T. H. (1986). Cultural-distance perspective: An exploratory analysis of its effect on learning and intelligence. *International Journal of Intercultural Relations*, **10**(4), 399–414.

Harrington, G. M. (1988). Two forms of minority-group test bias as psychometric artefacts with an animal model (*Rattus norvegicus*). *Journal of Comparative Psychology*, **102**(4), 400–407.

Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in Employment Testing: Validity Generalization and Minority Issues and the General Aptitude Test Battery*. Washington, DC: National Academy Press.

Hattrup, K., Schmitt, N., & Landis, R. S. (1992). Equivalence of constructs measured by job-specific and commercially available aptitude tests. *Journal of Applied Psychology*, **77**, 298–308.

Helms, J. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, **47**(9), 1083–1101.

Hickman, J. A., & Reynolds, C. R. (1986). Are race differences in mental test scores an artifact of psychometric methods? A test of Harrington's experimental model. *Journal of Special Education*, **20**(4), 409–430.

Hoffman, C. C., & Thornton III, G. C. (1997). Examining selection utility where competing predictors differ in adverse impact. *Personnel Psychology*, **50**, 455–470.

Holland, P. W., & Wainer, H. (eds) (1993). *Differential Item Functioning*. Mahwah, NJ: Lawrence Erlbaum.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, **9**(1/2), 152–194.

Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, **83**(2), 179–189.

Humphreys, L. G. (1992). Commentary: What both critics and users of ability tests need to know. *Psychological Science*, **3**(5), 271–274.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, **96**, 72–98.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, **86**, 721–735.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. M. (1977). Fairness of psychological tests: Implications of four definitions for selection utility and minority hiring. *Journal of Applied Psychology*, **62**, 245–260.

Jaradat, D., & Sawaged, S. (1986). The subset selection technique for multiple choice tests: An empirical inquiry. *Journal of Educational Measurement*, **23**, 369–376.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Johnson, J. W., Carter, G. W., Davison, H. K., & Oliver, D. H. (2001). A synthetic validity approach to testing differential prediction hypotheses. *Journal of Applied Psychology*, **86**(4), 774–780.

Johnson, S. T., & Wallace, M. B. (1989). Characteristics of SAT quantitative items showing improvement after coaching among Black students from low income families: An exploratory study. *Journal of Educational Measurement*, **26**, 133–145.

Jones, J. A., & Raju, N. S. (2000). Differential item and test functioning and cut-off scores in personnel decision making. Paper presented at the 15th Annual Meeting of the Society for Industrial and Organizational Psychology, New Orleans.

Kadlec, H. (1997). The correlation between the mean difference vector and first principal component holds even with discrete variables containing no *g*. *Cahiers de Psychologie Cognitive*, **16**(6), 750–757.

Kaufman, A. S., & Wang, J. (1992). Gender, race and education differences on the K-BIT at ages 4–90 years. *Journal of Psychoeducational Assessment*, **10**(3), 219–229.

Kober, N. (2001). *It Takes More than Testing. Closing the Achievement Gap*. Washington, DC: Center on Education Policy.

Kurz, R. (2000). The facets of occupational testing: general reasoning ability, residual attitudes and speed-accuracy balance. Unpublished PhD thesis, UMIST.

Kurz, R., Lodh, B., & Bartram, D. (1993). Practice and performance—the effects of systematic test preparation. Paper presented at the British Psychological Society Occupational Psychology Conference, Brighton.

Llabre, M. M. (1991). Time as a factor in the cognitive test performance of Latino college students. In J. Deneen, G. Keller, & R. Magallan (eds), *Assessment and Access: Hispanics in Higher Education* (pp. 95–104). New York: SUNY Press.

Llabre, M. M., & Froman, T. W. (1987). Allocation of time to test items: A study of ethnic differences. *Journal of Experimental Education*, **55**, 137–140.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race Differences in Intelligence*. New York: Freeman.

Lynn, R. (1996). Racial and ethnic differences in intelligence in the United States on the differential ability scale. *Personality and Individual Differences*, **20**, 271–273.

Mains-Smith, N. E., & Abram, C. M. (2000). *Evaluation of Item Bias and Test-taking Attitudes concerning the Royal Navy Recruiting Test (RT)* (DERA/CHS/MID/CR00547/1.0). Farnborough, UK: Defence Evaluation and Research Agency.

McLellan, R., Inkson, K., Dakin, S., Dewew, P., & Elkin, G. (1987). *People and Enterprises: Human Behaviour in New Zealand Organizations*. Sydney: Holt, Rinehart & Winston.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, **66**(4), 337–344.

Neisser, U., Boodoo, G., Bouchard, T. J. Jr., Wade Boykin, W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, **51**(2), 77–101.

Nyborg, H., & Jensen, A. R. (2000). Black-White differences on various psychometric tests: Spearman's hypothesis tested on American armed services veterans. *Personality and Individual Differences*, **28**(3), 593–599.

Nyfield, G., Gibbons, P. J., Baron, H., & Robertson, I. T. (1995). The cross cultural validity of management assessment methods. Paper presented at the 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL, May.

Office of National Statistics (2001). *Population by Ethnic Group and Age, Social Trends Data Set* (Social Trends 31, 2001 edn). London: The Stationery Office.

Oliver, M. L., & Shapiro, T. M. (1995). *Black Wealth/White Wealth: A New Perspective on Racial Inequity*. New York: Routledge.

Ones, D. S., Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology*, **75**(3), 255–276.

Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and demonstration of multi-dimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement*, **34**, 253–272.

Overton, R. C., Harms, H. J., Taylor, L. R., & Zickar, M. (1997). Adapting to adaptive testing. *Personnel Psychology*, **50**, 171–185.

Payne, B. D., Smith, J. E., & Payne, D. A. (1983). Sex and ethnic differences in relationships of test anxiety to performance in science examinations by fourth and eighth grade students—Implications for valid interpretations of achievement test scores. *Educational and Psychological Measurement*, **43**, 267–270.

Pennock-Roman, M. (1992). Interpreting test performance in selective admissions for Hispanic students. In K. F. Geisinger (ed.), *Psychological Testing of Hispanics* (pp. 99–135). Washington, DC: American Psychological Association.

Powers, D. E. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice*, **12**, 24–30.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, **9**(3), 241–258.

Pulakos, E. D., Schmitt, N., & Chan, D. (1996). Models of job performance ratings: An examination of ratee race, ratee gender, and rater level effects. *Human Performance*, **9**, 103–119.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measure of differential item functioning of items and tests. *Applied Psychological Measurement*, **19**, 353–368.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1993). *Student Group Differences in Predicting College Grades: Sex, Language and Ethnic Groups* (College Board Report No. 93-1). New York: College Board.

Rhine, W. R., & Spaner, S. D. (1983). The structure of evaluative anxiety among children differing in socio-economic status, ethnicity, and sex. *Journal of Psychology*, **115**, 145–158.

Richmann-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, **85**(6), 880–887.

Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, **55**, 171–183.

Robertson, I. T., & Kinder, A. (1993). Personality and job competencies: The criterion related validity of some personality variables. *Journal of Occupational and Organizational Psychology*, **66**, 225–244.

Rodriguez, O. (1992). Introduction to the technical and societal issues in the psychological testing of Hispanics. In K. F. Geisinger (ed.), *Psychological Testing of Hispanics* (pp. 167–172). Washington, DC: American Psychological Association.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer III, F., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, **54**, 297–330.

Roth, P. L., & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology*, **85**(3), 399–406.

Ryan, A. M. (2001). Explaining the Black–White test score gap: The role of test perceptions. *Human Performance*, **14**, 48–75.

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, **26**(3), 565–606.
Ryan, A. M., Ployhart, R. E., & Friedel, L. A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology*, **83**, 298–307.
Ryan, A. M., Ployhart, R. E., Greguras, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: Self selection and program effectiveness. *Personnel Psychology*, **51**, 599–622.
Ryan, A.M., Sacco, J.M., McFarlane, L.A and Kriska, S.D. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology*, **85**, 163–179.
Ryan, J. A., Schmidt, D. B., & Schmitt, N. (1999). Candidate orientation programs: Effects on test scores and adverse impact. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
Sackett, P. R. (1998). Performance assessment in educational and professional certification: Lessons for personnel selection. In M. D. Hakel (ed.), *Beyond Multiple-choice: Evaluating Alternatives to Traditional Testing for Selection* (pp. 113–129). Mahwah, NJ: Lawrence Erlbaum.
Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection. In C. L. Cooper, & I. T. Robertson (eds), *International Review of Industrial and Organizational Psychology* (pp. 146–185). New York: John Wiley & Sons.
Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, **50**, 707–721.
Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, **56**(4), 302–318.
Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of adjustment in pre-employment testing. *American Psychologist*, **49**, 929–954.
Salgado, J. F., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance*, **15**(1&2), 75–96.
Salmon, K. (1990). *The Use of Psychometric Tests in Personnel Selection and Recruitment and Their Implications for EEO Target Groups*. Wellington, New Zealand: Equal Employment Opportunities Unit, State Services Commission and Ministry of Women's Affairs, New Zealand.
Samuda, R. J. (1975). *Psychological Testing of American Ethnic Minorities*. New York: Harper & Row.
Scheunemann, J., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, **27**, 109–131.
Schiele, J. H. (1991). An epistemological perspective on intelligence assessment among African American children. *Journal of Black Psychology*, **17**(2), 23–36.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.
Schmidt, F. L., Pearlman, K. H., & Hunter, J.E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans. A review and analysis. *Personnel Psychology*, **33**, 705–724.
Schmit, M. J., and Ryan, A. M. (1997). Applicant withdrawal: The role of test-taking attitudes and racial differences. *Personnel Psychology*, **50**, 855–876.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. In C. L. Cooper, & I. T. Robertson (eds), *International Review of Industrial and Organizational Psychology*. New York: John Wiley & Sons.

Schmitt, N., & Mills, A. E. (2001). Traditional tests and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology*, **86**(3), 451–458.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, **82**(5), 719–730.

Schmitt, N., Sacco, J. M., Ramey, S., Ramey, C., & Chan, D. (1999). Parental employment, school climate, and children's academic and social development. *Journal of Applied Psychology*, **84**(5), 737–753.

Schoenemann, P. (1997). Famous artefacts: Spearman's hypothesis. *Cahiers de Psychologie Cognitive*, **16**(6), 665–694.

Shih, M., Pittinsky, T. l., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, **10**, 80–83.

SHL Group (2002). Internal database. Thames Ditton, UK: SHL Group.

SHL Group (1999) *OPQ32 Manual and User's Guide*. Thames Ditton, UK: SHL Group.

Steele, C. M. (1997). A threat in the air—How stereotypes shape intellectual ability and performance. *American Psychologist*, **52**(6), 613–629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, **69**, 797–811.

te Nijenhuis, J., & van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology*, **82**(5), 675–687.

Toplis, J., Dulewicz, V., & Fletcher, C. (1997). *Psychological Testing: A Manager's Guide*. London: Chartered Institute of Personnel and Development.

Trompenaars, F., & Hampden-Turner, C. (1993). *Riding the Waves of Culture*. London: Nicholas Brealey.

Verive, J. M., & McDaniel, M. A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. *Intelligence*, **23**, 15–32.

Vey, M. A., Hezlett, S. A., Kuncel, N. R., Ahart, A. M., Ones, D. S., Campbell, J. P., & Camara, W. J. (2001). The predictive validity of the SAT for racial and gender groups: A meta-analysis. Paper presented at the 2001 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, **52**(3), 679–700.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, **91**, 461–481.

Wild, C. L., Durso, R., & Rubin, D. B. (1982). Effects of increased test-taking time on test scores by ethnic group, years out of school and sex. *Journal of Educational Measurement*, **19**, 19–28.

Wing, H. (1980). Profiles of cognitive ability of different racial/ethnic and sex groups on a multiple abilities test battery. *Journal of Applied Psychology*, **65**, 289–298.

Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalisability across jobs and performance factors. *Personnel Psychology*, **43**, 355–366.

Wood, R., & Baron, H. (1993). *Towards Fair Selection. A Survey of Test Practice, and Thirteen Case Studies*. London: Commission for Racial Equality.

Chapter 7

IMPLICIT KNOWLEDGE AND EXPERIENCE IN WORK AND ORGANIZATIONS

André Büssing and Britta Herbig
Technical University München

INTRODUCTION

Over the last decades different areas of psychology have been increasingly researching the phenomenon of implicit knowledge—cognitive, work, and pedagogical psychology. Roughly speaking, cognitive psychology deals with the fundamental structure and processes of implicit knowledge, work psychology tries to explore the special achievements of implicit knowledge in working, and pedagogical psychology is mainly concerned with questions of knowledge imparting. Moreover, the question of the management of implicit knowledge and the use of this type of knowledge for expert systems is of great concern for both old and new economy organizations.

Although a huge bulk of research was conducted two main problems remain: first, up to now no consistent definition of implicit or tacit knowledge or even a uniform description of the phenomenon can be found. And, second, due to methodological issues there is only small or no transfer between different research directions or areas of application. The aim of this review is therefore twofold. On the one hand, we will try to give an overview on the different areas of research and their respective findings. Regarding the diversity of approaches and results this overview cannot be comprehensive and therefore aims at highlighting the most prominent research directions. And, on the other hand, we will try to evaluate and integrate the different results into a definition and research method that might be fruitful for theory as well as for application of implicit knowledge.

The review starts with some examples of the phenomenon of implicit knowledge. The section 'Approaches to research', deals in more detail with the aforementioned approaches to research and in the section, 'Implicit

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson
 ISBN: 0-470-84703-4

knowledge: A refined definition and an integrative approach', research results will be integrated in a refined definition of implicit knowledge. Furthermore, an integrative approach to research is presented that might be fruitful in considering cognitive as well as work psychology aspects. Finally, directions for future research and implications for the management of implicit knowledge in organizations are discussed.

IMPLICIT KNOWLEDGE—THE PHENOMENON[1]

In the literature several instances and examples are cited that are more or less openly connected to implicit knowledge. To give an impression of the phenomenon we will briefly present some of these examples.

The first example is cited by Kirsner and Speelman (1998): according to a rumor a leading French cheese producer spent several million francs on the development of an expert system to determine the ripeness of camembert. The latest knowledge elicitation techniques were used to identify the type of information employed by the experts. From the experts' responses it was concluded that the critical procedure occurred when the experts squeezed the cheese and that the crucial variable involved the tension of the cheese surface or, possibly, the pressure required compressing the cheese. Subsequently, an automatic system for measuring the surface tension of the cheese was developed—and failed completely. That is, the ripeness measurements of the system were systematically different from those of the experts. Subsequent research demonstrated that the actual information used by the experts were olfactory cues, not the surface tension, and that the olfactory information was released when the experts pinched the cheese just enough to break it.

This example demonstrates two of the most commonly mentioned features of implicit knowledge—the difficulty to verbalize this knowledge and its relation to action. The experts were able to access explicit knowledge about their expertise, and to provide verbal reports based on that knowledge. However, the explicit knowledge they named provided false information about the process despite the fact that they were experts in the task itself. The information the experts actually used was not 'open' to review although they used it successfully for years. In this case the consequences of the problem to gain access to implicit knowledge were 'only' financial. One dare hardly think of the hazards that could be produced by this problem were it used, for example, in medical expert systems or the operation of power plants.

A second class of examples reveals a property of implicit knowledge that is especially important in the workplace—the use and integration of often

[1] Although in some literature the term 'tacit knowledge' is used we will continuously use 'implicit knowledge' throughout the chapter because implicit is a broader term that also comprises tacit.

diffuse, sensory information. Carus, Nogala, and Schulze (1992) and Martin (1995) report instances of this type from the domain of Computerized Numerical Control (CNC) lathes. CNC lathes are completely enclosed for reasons of occupational safety; that is, only very little information, like sound, can escape. Nevertheless, workers who worked with the same CNC lathe for a long time were able to tell if something was wrong inside the machine. They brought the lathe to an emergency stop before a breakage of the tools took place. Thereby they were able to prevent financial losses due to machine idleness. Asked about how they knew that there might be a problem most of them were only able to state that they had a hunch or a sensation that something might be wrong inside the machine. Nevertheless, some workers could tell that the noise from the machine was somehow different or the vibrations were altered or even that they already had a bad feeling about the 'touch' of the processed material. An outsider could not perceive these diffuse sensations and the specific information configuration was difficult to name by the worker. This example shows why in work psychology sensory information gained by experience is seen as an important aspect of implicit knowledge.

Nonaka (1994) presents an example that hints at a similar quality of implicit knowledge—the development of an innovative home bread-making machine by a large Japanese company. Although the development team had all the technical knowledge to build such a machine it was decided to send a member of the team as an apprentice to one of the best bakeries in order to learn how to make really delicious bread. While working with the head baker the team member noticed that he had a very particular method of stretching the dough while he kneaded it. This experience was shared with the development team and implemented into the bread-making machine. The machine became a great success.

In this example sensory information again plays an important part but, moreover, it describes how the actual experience is sometimes necessary to learn specific know-how that in turn may constitute implicit knowledge.

However, there are also examples that paint a different picture of the suitability of implicit knowledge (for an overview see Mandl & Gerstenmaier, 2000). These examples are mostly investigated within the realm of other phenomena and explained by processes other than the ones we are discussing here, but nevertheless implicit knowledge may also play an important role in these phenomena. At least in Western society everybody is aware of the risks of smoking, the consequences of an unhealthy lifestyle resulting in problems like heart disease, or the ways in which HIV can be contracted, but still—people do smoke, eat too much fat, and do have unprotected sex. The same holds true for environmental issues: people are concerned about the exploitation of nature, the destruction of the ozone layer, and the diminishing sources of drinking water on Earth, but again—people do not recycle natural resources as much as possible, do drive cars even when it is

not necessary, and do use drinking water in abundance. These puzzling discrepancies between knowledge and action have their basis in a different feature of implicit knowledge, that is, its acquisition by implicit learning and experience, which in turn means that most people are not aware of their implicit knowledge. Implicit learning can hereby be characterized as a non-conscious process in which knowledge is acquired without the intention to learn something. Moreover, it is assumed that implicit learning is not selective; that is, all contingencies between different stimuli are stored (for an overview on implicit learning see Seger, 1994).

The common feature of these phenomena seems to be the gap between the knowledge and the actual, subjective 'beliefs' that people might non-consciously hold. In the case of health hazards these beliefs may consist of a too low probability estimation regarding the risk of becoming ill; for environmental problems a resigned attitude may exist. In many cases these beliefs have their roots in personal experiences, like 'my grandfather smoked his whole life and was in good health all his life'. In cognitive contexts these subjective beliefs are investigated under the headline of judgement fallacies and biases under bounded rationality, as Simon (1955) called it. A good example is the so-called 'base rate fallacy'; that is, people draw inferences from a wrong or too small a set of instances (e.g., Macchi, 1997; Stanovich & West, 1998). The base rate fallacy might be overcome if people are given the correct data for frequency of occurrence (e.g., Girotto & Gonzales, 2001). Here, implicit knowledge gives an additional explanation: subjective beliefs rooted in personal experience or acquired by implicit learning are difficult to overcome since in most cases they cannot be accessed consciously. That is, people will not be able to correct their base rate since they are not aware of it.

This presents the other side of the coin: implicit knowledge that is inadequate for certain situations but is used because of a lack of awareness for changing this knowledge (see Büssing, Herbig, & Latzel, 2002a).[2] Implicit knowledge has to be differentiated from inert knowledge, which is also used to explain the described phenomena. Inert knowledge means knowledge that can be explicitly stated, i.e., a person is aware of this knowledge, but it is not put into action. Several explanations for inert knowledge are possible, namely, meta-cognitive, structural deficit, and situativity explanations (see Renkl, 1996; Renkl, Mandl, & Gruber, 1996).

To sum up, the above-presented examples give a first glimpse at the phenomenon of implicit knowledge. Although they are not complete nor are they shared by all researchers or research directions, they comprise

[2] Nevertheless, because of limited knowledge, time, etc. rationality alone is not the best way of making decisions (e.g., Todd & Gigerenzer, 2000) and taking subsequent action; therefore, the violation of rationality as stated in the base rate fallacy might in some cases even lead to more accurate predictions in social situations than the use of rationality (Wright & Drinkwater, 1997).

some of the most often mentioned features of implicit knowledge. That is, the difficulty to put this type of knowledge into words due to a lack of consciousness; the problem that implicit knowledge might contain erroneous or naive theories; the importance of sensory information in implicit knowledge and its acquisition through concrete experience. In the following sections we will describe these features as well as divergent research findings in more detail.

APPROACHES TO RESEARCH

Implicit Knowledge in Basic Research

Although the effects of implicit knowledge are most conspicuous in applied contexts, the fundamentals for describing and understanding the phenomenon are based on research into implicit learning and knowledge. Therefore, an overview of this research will be given before looking at implicit knowledge in an applied context.

Research in cognitive psychology

Most research into implicit knowledge has been and is conducted within cognitive psychology. The experimental paradigms of cognitive psychology mostly contain tasks like serial reaction time tasks (e.g., Nissen & Bullemer, 1987; Willingham, Nissen, & Bullemer, 1989), artificial grammars (e.g., Reber, 1976, 1989), or control of dynamic systems (e.g., Berry & Broadbent, 1988; Broadbent, FitzGerald, & Broadbent, 1986; for an overview see Seger, 1994). The common denominator of these tasks is the implicit learning of artificial rules that have no relation to knowledge from 'real' life in order to ensure comparability of the knowledge bases of the test persons. As a vast amount of research has been conducted within these paradigms (for an overview see, e.g., Berry, 1997; Kirsner et al., 1998), the following overview on findings will be grouped according to the individual features of implicit knowledge.

An often-mentioned feature of implicit structures and processes is that they operate outside *consciousness* while explicit knowledge is always accessible to consciousness. However, this commonly used concept for contrasting the two modes of knowledge is not without problems. As O'Brien-Malone and Maybery (1998) point out, the concept of consciousness is by no means a homogeneous or coherent whole (e.g., Natsoulas, 1978, was able to identify at least seven different meanings of this concept). Two basically different points of view can be found regarding implicit knowledge and consciousness (e.g., Berry, 1997). Both positions assume that implicit learning is an unconscious process, i.e., there is neither consciousness for the learning process nor has the learner an intention to learn. The 'no-access' position (e.g., Lewicki, Czyzewska, & Hill, 1997), moreover, claims that this unconsciously acquired knowledge remains inaccessible to consciousness, while the 'possible-access'

position (e.g., Reber, 1989) claims that implicitly learned knowledge does not necessarily remain unconscious, i.e., implicit knowledge may be accessible to the consciousness. Looking at the findings from the research paradigms of cognitive psychology the 'no-access' position can hardly be maintained. By controlling the effects of implicit learning, participants were asked to name the underlying grammar rules (Reber, 1989), models of complex systems (Sanderson, 1989), or pattern rules (Hartman, Knopman & Nissen, 1989). At least some of the participants in the various studies were able to verbalize partially correct rules. Therefore, there is evidence against the 'no-access' position (for an overview see Shanks & St. John, 1994) and for the appropriateness of the 'possible-access' position except for one important limitation: although most participants performed much better after the implicit learning phase, the verbalization of assumed rules was rarely complete or completely correct. That is, implicit knowledge is not entirely unconscious but those aspects that are explicable do not reflect the whole, implicitly acquired knowledge about a task.

Dienes and Berry (1997) argue differently but with similar results. They state that implicit knowledge works below a subjective threshold; that is, implicit knowledge is not consciously perceived as guiding one's actions. The focus of this argument is therefore not the acquisition but the use of implicit knowledge and a more specific definition of the role of consciousness. The importance of implicit knowledge for the guidance of actions clearly does not prohibit the possibility of awareness of this knowledge at other times nor does it propose that a conscious engagement in some kind of action is impossible if this knowledge type is used. Working activities that are guided by experience are especially subject to involvement of implicit knowledge in this sense. A subjective threshold is defined by the level of discriminatory answers for which persons state that they no longer detect perceptual information, that is, that they are just guessing, although they perform at an above-chance level (e.g., Cheesman & Merikle, 1984). Knowledge above a subjective threshold is conscious and can be defined as explicit knowledge. Dienes and Berry (1997) point out that the assumption of a subjective threshold may explain and integrate findings from different research areas. One interesting result in this context is that, regardless of the subjective threshold, people do have a kind of rudimentary meta-knowledge of their implicit knowledge. That is, by questioning people about how much they trust their answers in implicit tasks, they showed a higher trust in correct answers than in incorrect ones, although they claimed that they were just guessing (Chan, 1992). Therefore, the statement that implicit knowledge is not consciously perceived as guiding one's actions does not prohibit the possibility of awareness of this knowledge at other times. The trust in one's own knowledge and ability is also an important aspect of experience-guided working and is therefore a common denominator in cognitive and work psychology.

Closely related concepts to consciousness are *awareness* and *intention* in implicit knowledge. Implicit knowledge is believed to work without intention and awareness while explicit knowledge cannot be acquired or used without consciousness, awareness, or intent. At least theoretically, there seems to be little doubt that implicit knowledge may occur when there are no conscious, reflective strategies to learn (Reber, 1989); that is, that the acquisition of implicit knowledge can happen incidentally. Empirically, this assumption is difficult to test since in the research paradigms of implicit learning the stimuli are nearly always at the forefront of a participant's attention. Only a few investigations have tried to bring about implicit learning under conditions of minimal attention, like, for example, research into implicit perception in which awareness that something should be learned is prevented by attention manipulation (e.g., MacLeod, 1998). Results indicate that for some types of tasks a mere exposure effect is sufficient in order to learn relations between stimuli whereas other types of task need a higher degree of attention (e.g., Greenwald, 1992). Therefore, there is no conclusive answer to the question about the necessity of awareness for the acquisition and use of implicit knowledge.

Another important question in cognitive research concerns the *complexity* of implicit knowledge. For knowledge to be termed complex it has to comprise a great number of elements that have manifold connections among them. Undoubtedly, explicit knowledge can be complex in this way (see, e.g., Preussler, 1998), but for implicit knowledge contradictory research results have been found. For example, on the one hand, social cognition research shows that people are not able to name proportion rules for human faces but react to even the slightest aberration from these complex rules (e.g., Lewicki, 1986). On the other hand, computer simulations of artificial grammar research imply that participants in this research did not always learn the complex rules but that results can also be explained by the learning of simple letter pairs (e.g., Ericsson & Smith, 1991). These examples show—even for quite simple contexts—the divergence between the results and therefore the difficulty to give a final evaluation of the complexity of implicit knowledge. At least theoretically, implicit experiential knowledge, as described in the section, 'Research in developmental and pedagogical psychology', should be more complex than the knowledge investigated in cognitive psychology (Mathews, 1997).

Flexibility of knowledge means being able not only to transfer knowledge to different situations and areas but also to be able to combine and link different parts of knowledge. Both complexity and flexibility are often viewed together whereby it is regularly assumed that consciousness and therefore explicit knowledge is a precondition for flexibility respectively for the flexible use of knowledge (Browne, 1997). The inverse assumption is that implicit knowledge itself is not flexible. Holyoak and Spellman (1993) characterize implicit knowledge as a complex structure but at the same time

assume that it is inflexible and therefore difficult to transfer. As a possible reason for this lack of flexibility Gazzard (1994) names a simultaneous, one-dimensional processing of stimuli. Explicit knowledge, on the other hand, allows for a simultaneous, multi-dimensional processing that is more flexible. An indication for the correctness of this assumption is given in an investigation by Willingham et al. (1989). In serial reaction time tasks, implicit learning was established by an above-chance detection of the underlying pattern. Nevertheless, participants who could name the pattern reacted faster in subsequent performance than persons who 'only' applied implicit knowledge. A transfer of implicit knowledge into an explicit mode might therefore be a necessary precondition for flexible use. Unfortunately, cognitive research paradigms do not investigate the explication of implicit knowledge as an active process. Rather, the only theory on this problem, by Karmiloff-Smith (1990), declares that a sufficient amount of implicit knowledge has to be acquired so that this knowledge becomes explicit. This automatic process is called representational re-description; that is, well-learned and repeated implicit representations are subjected again and again to renewed descriptions until the knowledge structures show a higher flexibility and are accessible to consciousness and verbalization.

Research in developmental and pedagogical psychology

Developmental psychology mostly investigates implicit knowledge within the cognitive development of children. The basic assumption of this research is that implicit knowledge developed evolutionarily before 'higher' cognitive processes ('primacy of the implicit', Reber, 1993) and therefore contains a type of naive theory on the world and its connections (Macrae & Bodenhausen, 2000; Olson & Campbell, 1994). In the course of development this knowledge should then be replaced by theories that are more adequate (Weinert & Waldmann, 1988). Some developmental psychology research findings show that this replacement does not take place. That is, implicit, naive theories persevere independently alongside explicit theories and gain the upper hand in certain situations (e.g., Fischbein, 1994; Gelman, 1994; Sternberg, 1995). As it is assumed in developmental and pedagogical psychology that these implicit theories are mostly incorrect, this perseverance is seen as a deficit that has to be overcome (e.g., Clement, 1994; Lee & Gelman, 1993). Moreover, it was shown that even those persons who were provided with plenty of evidence against their implicit theories continued to use these theories especially in difficult situations. That is, implicit knowledge is very resistant to change even if opposing explicit knowledge does exist (Weinert & Waldmann, 1988).

Two different approaches can be observed in pedagogical psychology. The first one is similar to developmental psychology and defines implicit knowl-

edge as naive, sometimes erroneous theories about the world that have to be transformed into more adequate representations. The second approach is anchored in Reber's theory (1993), which also assumes the 'primacy of the implicit' (1993) but without accepting at the same time that implicit knowledge is inferior. Here, implicit processes are embedded in an evolutionary viewpoint that claims that consciousness developed relatively late in evolution meanwhile sophisticated, unconscious, perceptive, and cognitive functions existed a long time before its development. In this approach complex knowledge that is acquired during an implicit learning task is represented in a general, abstract form. This form only contains little information on the specific stimuli configuration but stores the structural relations between the stimuli. Abstract representations in the context of non-consciousness are strongly debated (e.g., Neal & Hesketh, 1997) as the ability for abstraction is solely attributed to 'higher' and, in this argument therefore, conscious cognitive processes. Reber (1993), on the other hand, argues that in an extreme, complex environment the ability for abstraction has a high adaptive value and, because of that, should have developed very early on in phylogenesis. Research on this question—commonly conducted with patients who have experienced neurological insult or injury—is inconclusive (for overviews see Schacter, McAndrews, & Moscovitch, 1988; Shimamura, 1989).

Nevertheless, with this change in perspective from basic 'problem implicit knowledge' to 'chance implicit knowledge', tendencies in pedagogical psychology can be found to use implicit knowledge in a constructive way. Catchwords for these tendencies are 'learning by doing', 'learning by osmosis', 'professional instinct', or 'intuition'. Moreover, Macrae and Bodenhausen (2000) point out the importance of implicit theories for social cognition. Implicit models do have a great influence on our cognitive processes (Fischbein, 1994), and this influence is rooted most probably in its empirical origin. Implicit models correspond with our experience, while theoretical interpretations are based in logical coherence. Therefore, at least under certain circumstances, empirical-based models do have a greater impact on our thinking than conceptual models. With this notion experience is introduced as an important factor of implicit knowledge in pedagogical psychology. Consequently, the possibility for concrete experience is seen as an essential part of knowledge acquisition. This assumption is closely related to concepts from applied psychology.

Implicit Knowledge in Applied Psychology

Although different theories on implicit knowledge in organizations exist (e.g., Baumard, 1999; Dierkes, Antal, Child, & Nonaka, 2001; Nonaka & Takeuchi, 1995; Sternberg & Horvath, 1999) most of these theories are based on the work of Polanyi (1962, 1966) who was the first to describe

this phenomenon. In the section below, we will first give an overview of Polanyi's theory before describing its use in today's organizations. We will adopt Polanyi's use of the term 'tacit knowledge' in the next section.

Theoretical foundations—Polanyi's theory on tacit knowledge

Polanyi's most basic notion on implicit or tacit knowledge is 'we can know more than we can tell' (Polanyi, 1966, p. 4), which describes the fact that implicit knowledge is not easily verbalized and therefore is difficult to exchange between individuals. The reason for this is that tacit knowledge is entrained in action or practice and linked to concrete contexts. Tacit knowledge is developed through concrete sensory experience and the integration of various impressions into a holistic picture of a situation. Based on findings from Gestalt psychology Polanyi views the 'Gestalt' as the result of the active molding of experience during the process of realization. In order to develop tacit knowledge people have to empathize with the objects of the world; they have to take them in. For a learner to be successful in this intake, he or she has to assume that what needs to be learnt is meaningful even if it seems senseless at the beginning. Through different steps of integration and interpretation, seemingly meaningless sensations and/or feelings are translated into meaningful ones and transformed into experience. For example, when using a tool the degree of pressure on the hand is registered and controlled by the effect made on the object. Therefore, implicit learning brings about a meaningful relation between different aspects of a situation. Polanyi (1962) describes this learning as the understanding of complex entities by means of which bodily and sensory perceptions play an essential role. Implicit knowledge and learning can therefore be seen as empathy; that is, people will not understand complex entities by looking at things, but will do so only through empathy. This process of empathic understanding depends on a kind of perception in which information is seen in terms of the whole. The structure of implicit knowledge consists of a from–to relation between information, parts, characteristics, and the focal whole, to which they are related by the mental act of integration. This constitutes the functional reference between the two poles of the 'tacit relation'—the details, on the one hand, and the focal whole, on the other (Polanyi, 1966; Sanders, 1988). Experience, as the implicit construction of the world, depends on two processes—integration and differentiation. Implicit knowledge as the perception of entities is based on the integrative function but, moreover, Polanyi (1966) describes the perception of details and their differences as the most important task for a deeper understanding of these entities. Therefore, experience and thus implicit knowledge is built up by a constant change between integration and differentiation.

Implicit knowledge in work psychology

From the perspective of work psychology implicit knowledge is mostly investigated in the context of work experience and therefore as an essential part of experiential knowledge (e.g., Olivera, 2000). Exemplary catchwords for implicit knowledge are flair for a material or an intuitive grasp on intricate or difficult situations. This implicit knowledge is individually acquired by the worker through events in work situations. It is embedded in the working process; that is, it is learned implicitly in the course of action. Implicit knowledge is therefore bound to a person and is situation- or context-oriented. This so-called implicit experiential knowledge results in specific performance in a work situation that cannot be mastered solely by routine (Carus et al., 1992). A common characteristic of these types of work situation is that they are not completely describable in advance; that is, they cannot be standardized and rules are not sufficient for mastering these situations. Moreover, implicit experiential knowledge is of utmost importance in situations in which a great number of different interrelated process parameters have to be manipulated or optimized (e.g., Martin, 1995). Therefore, in contrast to findings from cognitive psychology (see the section on 'Research in cognitive psychology') implicit knowledge in work psychology from this perspective is seen as complex, multi-dimensional, and very flexible knowledge.

There are other concepts defined in work and pedagogical psychology that relate in some way to implicit knowledge. From the viewpoint of conceptualization of knowledge these concepts are 'situated knowledge' and the distinction between 'global and specific knowledge'. For the acquisition of knowledge, apprenticeship approaches and the model of experience-guided working can be distinguished.

Situated knowledge is defined as knowledge that is principally bound to a situation (Greeno, 1998; Menzies, 1998) and therefore to certain antecedence conditions in order to be put into action. That is, situated knowledge has a close relation to action only if the environmental surroundings are very similar to those in which the knowledge was learned (Greeno, Smith, & Moore, 1993). The same holds true for the verbalizability of situated knowledge; that is, situated knowledge can only be stated if very similar conditions to the acquisition situation are created. Hence, it shares a common denominator with implicit knowledge.

The distinction between global vs. specific knowledge (e.g., Doane, Sohn, & Schreiber, 1999; Higgins & Baumfield, 1998) describes the discrimination of general thinking and problem-solving skills (global knowledge), on the one hand, and domain-specific knowledge that is useful only for certain problems or situations, on the other hand. Implicit experiential knowledge as seen in work psychology is bound to certain situations and contexts and might therefore be seen as predominantly specific knowledge. Although, as Higgins and Baumfield (1998) state, in recent years global knowledge has been neglected

in favor of specific knowledge, global knowledge might be very important for transfer achievements. Since implicit knowledge in the work context is said to be of high flexibility (e.g., Büssing, Herbig, & Ewert, 2001) it may well be that implicit knowledge also contains global knowledge.

The theory of situated knowledge is closely related to the global vs. specific dimension in so far as situated knowledge seems to describe specific knowledge. Since, in both cases, successful transfer depends on structural invariance of the interaction between actor and situation the problem of flexibility of implicit knowledge is also an issue with situated knowledge.

Besides the explicit learning of professional knowledge, two theories that are more complementarily than mutually exclusive try to explain the acquisition of implicit knowledge in the work context—apprenticeship approaches and experience-guided working. Apprenticeship as a way to acquire implicit knowledge has been described in a multitude of contexts (e.g., Ainley & Rainbird, 1999; Hay & Barab, 2001; Rimann, Udris, & Weiss, 2000) and can be divided into cognitive apprenticeship and apprenticeship in practical skills. Collins, Brown, and Newman (1989), who were the first to introduce the cognitive apprenticeship approach, differentiated between easily explicated knowledge about facts and implicit strategic knowledge from expert practice. This implicit knowledge is difficult to explicate outside an authentic problem situation. Moreover, it is best imparted in a situated way as well as within the social exchange between experts. Models for this imparting of implicit knowledge are traditional crafts, which are mostly limited to the domain of manual skills. The cognitive apprenticeship approach tries to transfer the use-oriented principles of knowledge imparting to cognitive domains with complex problems.

Apprenticeships in practical skills start with an observation of the actions of an expert by the apprentice. In cognitive apprenticeship this has to be supplemented by a verbal report of the expert about his cognitive processes and strategies in dealing with an authentic problem. The following steps in the knowledge acquisition process are quite similar for apprenticeships in practical skills and cognitive apprenticeships. The learner gets the opportunity to deal with a problem/task by himself. Thereby, the expert supports him through coaching and scaffolding. These measures of support slowly fade according to the progress and experience of the learner. In using the acquired skills and knowledge for a variety of tasks and problems the learner himself gains implicit strategic knowledge in the respective domain.

Complementary to these apprenticeship approaches the model of experience-guided working (Carus et al., 1992) and the subconcepts of subjectifying and objectifying action (Böhle & Milkau, 1988) allow a description of the ongoing acquisition of implicit knowledge in working. Subjectifying and objectifying action are both part of experience-guided working; therefore, some distinctions between the two forms of actions need to be outlined first.

The reference points for action in subjectifying action lie in concrete and

unique qualities and variations, while in objectifying action universally valid and generalizable rules dominate. Whereas in subjectifying action emotions play an important role for the structuring of action, in objectifying action they are only subordinate or disturbing elements in the work process. As for the sensation of the actor in subjectifying action, perception happens by means of complex sensations and by movements of the whole body. Moreover, emotions are an essential part of perception. For example, a nurse who enters a sick room sees the patient, perceives the smell in the room, hears the patient's breathing, and in touching the patient might get information about the condition of his/her skin and pulse, etc. This mass of sensory information may lead to the feeling that something is wrong with the patient, and it is upon this feeling that the nurse acts.

In objectifying action, on the other hand, only single senses are employed for exact, objective perception and, again, emotions are viewed as disturbing for objective perception. Taking the same situation, a nurse who acts objectifyingly might only hear that the patient breathes shallowly and then uses a stethoscope to concentrate on hearing and get an exact measure. A heightened state because of the feeling that something is wrong with the patient is here seen as disturbing the concentration needed for measurement.

In subjectifying action the environment has subject characteristics, while in objectifying action it has object characteristics. Subjectifying action is of a dialogic–interactive nature and in it the simultaneity of action and reaction is experienced. In objectifying action, however, the environment is either influenced one-sidedly, or influences and information are taken up reactively from the environment. Goals as well as concrete procedures only develop during the course of subjectifying action, while the planning of action steps and goals precedes objectifying action. Both forms of action cannot be compensated for or replaced by each other since they achieve different things. In experience-guided working a mutual crossing and completion of these two forms is assumed. Figure 7.1 presents how the different forms of knowledge and action might be related (see Büssing, Herbig, & Latzel, 2002b).

Polanyi's theory on tacit knowledge (1962, 1966) as well as apprenticeship approaches (e.g., Ainley & Rainbird, 1999; Hay & Barab, 2001; Rimann et al., 2000) show that human experience is necessary when it comes to reacting flexibly and effectively in unpredictable, critical situations. Experience in the context of experience-guided working is not only seen as a precondition for and a product of action but also as a process that can produce new patterns and insights at the moment of realizing the gap between real and expected situational conditions (Carus et al., 1992).

This perspective highlights how an individual actively deals with conditions of the environment during the course of experience development. The notion that experience is bound to the action process therefore focuses not only on the subject of experience but also on the field of experience and the respective conditions of experience (Böhle & Milkau, 1988). For example,

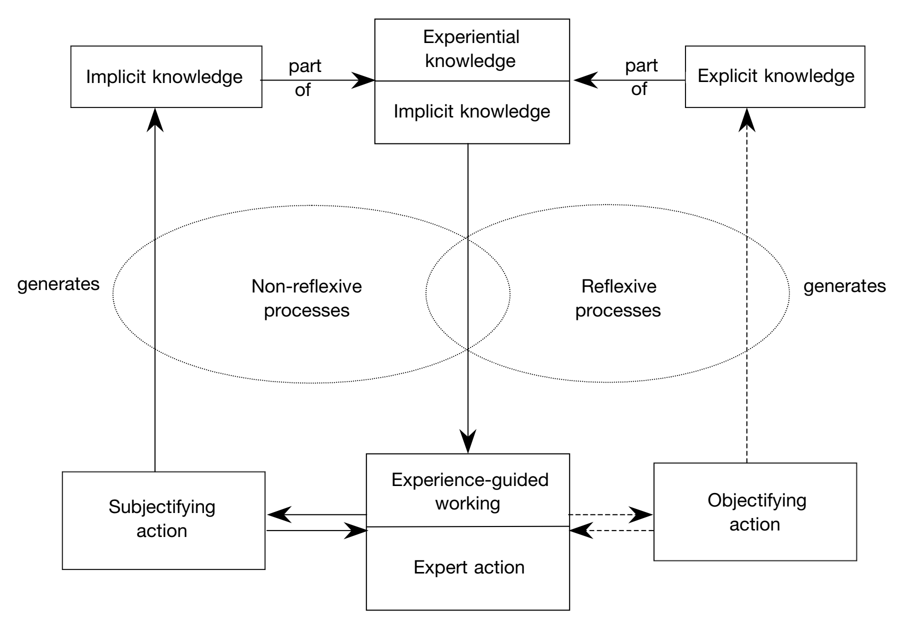


Figure 7.1 Model of implicit knowledge, experience, and action (MIKEA; reproduced by permission of Büssing et al., 2002b).

Wehner and Waibel (1996) showed, by simulating a critical situation in shipping, that experienced captains acted more rhythmically in the situation than inexperienced ones and were therefore able to handle the problem more efficiently with less stress than their colleagues.

Relying on this assumption the following characteristics of experience-guided working can be summarized:

- Complex sensory perception through several senses and perception of relations; senses are not particularized they are used simultaneously.
- Attention is distributed; that is, it is focused on symbols as well as on objects, processes, and movements.
- Perception of diffuse, not exact defined information.
- Vivid and associative thinking.
- No separation between planning and execution; pragmatic stepwise procedure and practical testing.
- Holistic images or patterns render the sequential–analytic interpretation of essential information partially unnecessary. Therefore, they allow for a time-critical development of strategies and evaluations of system states even in unpredictable, chaotic situations.

Subjectifying action and experience-guided working depend on the develop-

ment of holistic and flexible anticipation characteristics. That is, holistic mental images about what a situation should look like are then compared with the actual situation. They differ from situational images in objectifying action in so far as the comparison uses a similarity principle rather than an identity principle. This allows mental simulation with several diffuse variables. Therefore, in experience-guided working two different types of knowledge can be found: first, there are objective rules and exact information for use in the comparison for identity. This type of knowledge is codifiable, can therefore be reported, and seems to be rather explicit. Second, there is diffuse information in various combinations for use in the comparison for similarity. This type of knowledge might be difficult to report and seems to be rather implicit.

To sum up: from the perspective of work psychology implicit knowledge is an essential part of experiential knowledge and is individually acquired by a worker in the course of (holistic) working, which means implicit knowledge is bound to a person and is situation- or context-oriented. Therefore, an explicit imparting of this knowledge is hardly likely. Experience means the development of holistic and flexible anticipation characteristics, i.e., expectations and ideas about what a situation should look like. Anticipation characteristics are based on similarity principles, which allow mental simulations of the situation with a multitude of influencing variables (e.g., Büssing et al., 2001). Experience also includes the ability to use certain action patterns without becoming aware of their individual parts. Therefore, experience is not only a precondition and a product of action but also a process that produces a new 'Gestalt' at the moment of deviation between supposed and real conditions, and can lead to insight. This position highlights, on the one hand, how an individual actively deals with environmental conditions in the development of experience. On the other hand, it shows a close relation to the acquisition of implicit knowledge as conceptualized by Polanyi (1966).

Several conclusions about the features of implicit knowledge, as defined in work psychology, can be drawn. First, implicit knowledge can have a complex structure; that is, since implicit knowledge incorporates a variety of sensory and diffuse information in a multitude of combinations it is assumed that it must have a quite complex structure. Second, implicit knowledge can be flexible; that is, since mental simulations on the status of a work situation can be conducted, a transfer of implicit knowledge should be possible even to unknown situations. Third, implicit knowledge also contains, besides knowledge of facts and procedures, emotions and person-related knowledge as a consequence of the assumed acquisition mode. A more indirect conclusion is concerned with the question of the adequacy of implicit knowledge. The description of the use of implicit knowledge in work psychology mostly shows an adequate and sometimes highly intriguing use of this knowledge as a function of work experience. However, the question

about what happens if no experience with certain situations exists is not addressed (for an overview see Herbig, 2001; Herbig & Büssing, 2002).

These conclusions lead to a *fundamental problem* in the study of implicit knowledge in work psychology. The methods employed to investigate this type of knowledge are observation of people at work and interviewing them afterwards. Conclusions about implicit knowledge are drawn by inference from these observations and interviews. That means it is neither possible to determine whether the used knowledge was really implicit nor is it possible to understand the structures and contents of implicit knowledge in depth. Moreover, since the verbalizability of implicit knowledge is poor, interview data might contain erroneous or self-serving statements (e.g., Sharp, Cutler, & Penrod, 1988) that are not in line with real, employed knowledge.

Implicit knowledge in expertise

Only in the last few years has it become apparent that implicit knowledge may play an important role in expertise although the relation between experience and implicit knowledge was outlined by Polanyi as early as 1962. One of the problems might have been that a common definition of expertise is still lacking or, as Sloboda (1991) states, there is no consensus among experts on the issue of expertise. Although expertise is a fairly heterogeneous concept researchers agree that experts usually work faster, more precisely, and efficiently than novices and need less resources (Sonnentag, 2000; Speelman, 1998). Usually, experts do not display higher expenditures of energy or better physical abilities than novices. Differences between novices and experts are found above all in qualitative aspects—in the organization of performance prerequisites for a flexible, situation-, and goal-oriented use of resources as well as in meta-knowledge and strategies (Hacker, 1992). Action-guiding psychological images have a special position in this organization of performance prerequisites as they reduce the working memory load by compressing the knowledge. Thereby capacities to deal with complex characteristics of a situation are released. This concept of psychological images is very similar to 'holistic anticipation characteristics', a term used in experience-guided working. Both notions imply that implicit experiential knowledge achieves special results in working situations that cannot solely be accomplished by routine. Therefore, implicit knowledge seems to be an important component of expertise although the question about how implicit knowledge enhances performance in concrete working situations has yet to be answered. Nevertheless, there are some aspects that point to a relation between implicit knowledge and expertise because they show obvious parallels and connections.

One of the most prominent similarities between implicit knowledge and expertise and one of the biggest challenges for knowledge management in organizations can be found in the lack of verbalizability in both areas (e.g.,

Dreyfus & Dreyfus, 1986). This circumstance especially influences the development of expert systems in a negative way. A problematic factor in the elicitation of expert knowledge is that experts master rules in such a way that they recognize situations in which the use of the rules is not appropriate (Reber, 1993). This constitutes a serious restriction for the codification of expert knowledge and therefore a problem for the development of expert systems. Moreover, when experts report their knowledge the verbalized rules often do not match what actually happened. Similar results were found in research into implicit knowledge (Gazzard, 1994). Such incorrect or incomplete reports can cause immense problems when used in expert systems (e.g., financial losses in industry or health hazards in medicine).

Another similarity between implicit knowledge and expertise can be found in the form of acquisition. Expertise is mostly determined by the length of time spend in conducting a certain work or activity (Hacker, 1998). Expertise as well as implicit knowledge are generated mainly in concrete (working) situations; that is, they are not abstractly imparted but acquired through concrete actions in relevant contexts (Myers & Davis, 1993; Polanyi, 1962, 1966; Speelman, 1998). Experience-guided working therefore determines the acquisition of expertise.

There are models that allow implicit knowledge to be related to expertise theoretically (e.g., the Adoptive Control of Thought (ACT) model of Anderson, 1982, 1992; Speelman, 1998). This ACT model renders, at least in part, an explanation for the acquisition of implicit knowledge. Anderson describes, using this model, the acquisition of explicit knowledge that is transformed step by step into procedural knowledge. The resulting knowledge cannot be accessed easily and therefore has a common denominator with the definition of implicit knowledge. Nevertheless, research from the domain of medical diagnostics (Griffin, Schwartz, & Sofronoff, 1998) shows that the acquisition of implicit and explicit knowledge takes place in a parallel manner and that the findings cannot be explained completely by the ACT model. This holds especially true if expertise is differentiated into adaptive and routine expertise (Hantano & Inagaki, 1986). Adaptive expertise develops through activities in an area with different tasks and demands, and is easier to verbalize. Routine expertise, on the other hand, is more likely to develop in an area of activity that is essentially characterized by constancy. This type of expert knowledge is difficult to verbalize and is of limited flexibility. The development of routine expertise can be explained with the ACT model, but it is difficult to describe the development of adaptive expertise within the framework of the model (Speelman, 1998). Moreover, although there is no definitive statement in the model that compiled knowledge was formerly conscious and therefore explicit, Anderson (1982) only utilizes examples in which this is the case. That is, the ACT model is not able to explain the direct acquisition of implicit knowledge—knowledge that has never been conscious.

As for the mental representations in expertise and implicit knowledge, there are indications that knowledge representations that experts have in their domain differ from those in other domains or from those of novices (Rouse & Morris, 1986). Büssing et al. (2001) present first evidence that at least part of the knowledge representations of experts may indeed be implicit by nature.

The acquisition of implicit knowledge in working by non-reflexive processes (see Figure 7.1) highlights the central function of subjectifying action for this type of knowledge. Every new experience within a domain enlarges this knowledge. And, finally, it influences the action as implicit experiential knowledge, without the actor being aware of its action-guidance. Under the premise that this knowledge is adequate, performance can be enhanced. That is, with years of activity or working in a domain implicit knowledge is accumulated that forms part of the expertise of the person.

Implicit knowledge in organizational psychology

As presented above implicit knowledge as an essential part of expertise is a very important human resource when it comes to dealing with critical situations at work. Therefore, it is also of great concern for organizations to know how to manage this type of knowledge, especially since knowledge has become the most strategically important resource and competitive advantage for companies in advanced, information-driven societies (e.g., Drucker, 1993; Sveiby, 1997; Thurow, 1997). Therefore, organizational psychology mostly deals with knowledge management when considering implicit knowledge. Following an approach by Reinmann-Rothmeier and Mandl (2000), who divided knowledge management into four interlinked, yet distinctive process categories, we will present the role played by implicit knowledge in each of these categories. The categories are namely knowledge generation, knowledge representation, knowledge communication and knowledge use.

Knowledge generation comprises all processes used to obtain knowledge. This contains external knowledge generation (e.g., new employment, cooperation, or fusion) or the set-up of special knowledge resources within the organization (e.g., development departments); that is, a combination of explicit knowledge (see Figure 7.2). Moreover, Nonaka and Takeuchi (1995) (again see Figure 7.2) describe internalization (i.e., the transition of explicit to implicit knowledge through continuous use) and socialization (i.e., the non-conscious adoption of rules, views, etc. through social interaction) as ways of knowledge transition in organizations. Besides these types of knowledge transition and generation, the 'externalization' of knowledge plays an increasingly important role as a means of knowledge generation because it is the key to a hardly duplicable generation of knowledge. The term 'externalization' already hints at the fact that here implicit (i.e., personal), context-specific, and difficult to verbalize knowledge should be transformed into

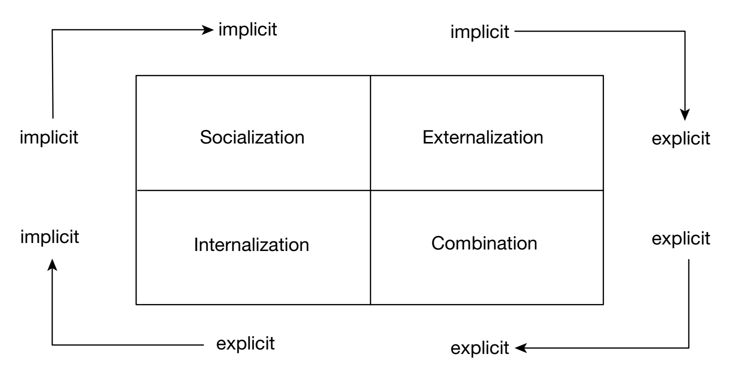


Figure 7.2 Transitions between explicit and implicit knowledge (data from Nonaka & Takeuchi, 1995).

explicit communicable knowledge. This communicability is not necessarily a verbal one. One might assume that externalization of an action itself could be difficult in a verbal mode (e.g: 'How do I drive a car?'), meanwhile the verbal externalization, thus explication, of action-guiding knowledge should be possible. Generally in the literature on knowledge management and organizational psychology, three different ways of externalization are described: apprenticeship, working in groups, and the development of expert systems.

Apprenticeship, as described in the section on 'Implicit knowledge in work psychology', involves the individual learning of a task by doing it under the supervision of an expert. Thereby, implicit knowledge is transferred from one individual to another without the necessity to explicate this knowledge completely (e.g., Cimino, 1999; Nonaka & Takeuchi, 1995; Polanyi, 1966). The processes involved in apprenticeship as implicit learning are closely related to experiencing and therefore to Polanyi's description of tacit knowing (1962) and/or experience-guided working (e.g., Martin, 1995). Nonaka and Takeuchi (1995) call this process 'socialization' to stress the transformation from implicit knowledge in one person to implicit knowledge in another person (see Figure 7.2).

Another way of externalization is presented by working in groups, that is, mostly people working together who have all types of specialized skills (e.g., Johannessen & Hauan, 1994; Johannessen, Olaisen, & Hauan, 1993). The central process assumed to be of importance to knowledge generation in groups is a learning loop where continuous improvements, by learning through doing, using, experimenting, and interacting, create a positive spiral for innovation (e.g., Johannessen, Olaisen, & Olsen, 2001). In this case, not only the shared reality of work experience, which again is a type of socialization, is of importance but also the communication of implicit knowledge. A prototypical way to communicate implicit knowledge, although it is difficult to verbalize, is so-called 'storytelling'. Storytelling as a method (e.g., Roth & Kleiner, 1998; Swap, Leonard, Shields, & Abrams, 2001)

comprises interviews on important recent events in an organization, the extraction of themes, and a newly organized story of the history of the organization, which is then validated and used as an starting point for further communication between the members of an organization. But storytelling is also an important aspect of sharing implicit knowledge within work groups. Here, the narration of experiences from past events serves to make these experiences transparent without the necessity to verbalize the gained knowledge in a structured and comprehensive way (e.g., Baumard, 1999; Benner, 1984).

The third way of externalization of implicit knowledge is the use of knowledge elicitation systems (for an overview see Firlej & Hellens, 1991; Lant & Shapira, 2001) in order to gain implicit and explicit expert knowledge. This expert knowledge is then communicated to others mostly via information systems in the form of expert systems (for an overview see, e.g., Darlington, 2000; Jackson, 1999; Liebowitz, 1998). While apprenticeship comprises the transfer of knowledge from one person to another person and work groups comprise the transfer from several persons to other persons, expert systems allow the transfer of knowledge from one person or a small group of experts to a large number of persons. However, this type of externalization is especially problematic since research into expertise showed the difficulty experts experience in verbalizing their knowledge at all or correctly (e.g., Ericsson & Smith, 1991, see also the section on 'Implicit knowledge in expertise'). Although this is not a problem that is unique to experts, it may be of special interest here. In apprenticeship and work groups the direct interaction gives (implicit or explicit) feedback on the adequateness of the verbalization, while with expert systems users are separated in time and space from the expert(s). That is, feedback or more comprehensive explanations are not possible.

Therefore, an examination of the externalized knowledge for expert systems needs to take place. Moreover, Herbig (2001) was able to show that the reintegration of externalized knowledge within the knowledge recipient may cause problems if no opportunity is given to use and therefore experience this knowledge.

Although this reintegration problem does not concern knowledge generation via apprenticeship or work groups, Baumard (1999) describes another difficulty within these types that is concerned with motivational and social psychology questions. People holding implicit knowledge have to be prepared to impart their knowledge and to use it in a constructive way. Work groups, for example, may use their implicit knowledge to create so-called 'fuzzy zones' around them in order to demarcate them from others. This in turn would undermine every attempt for knowledge management in an organization.

Knowledge representation is another part of knowledge management that describes all the processes of codification, documentation, and storage of

knowledge. The problem of implicit knowledge in this area is a quite straightforward one: implicit knowledge is by definition knowledge that is difficult or even impossible to codify and therefore cannot be documented in an adequate way (e.g., Ryle, 1993; Sproull & Kiesler, 1994). Moreover, concerns have been voiced that, with a growing formalization of knowledge, implicit knowledge may fade away and that this resource might be lost for organizations (e.g., Campbell, 1990; Johannessen et al., 2001). Nevertheless, given the assumption that at least some implicit knowledge can be made explicit a few implications for knowledge representation exist. For organizations implicit knowledge is of concern as described in work psychology, that is, complex and flexible knowledge that connects a multitude of different information and sensations in a meaningful way. Therefore, databases that represent this type of knowledge have to be structured in a highly connected way. In recent years the development of databases has tried to implement such a connectivity in a user-friendly way (e.g., Kriegel, 2000) but this does not guarantee the acquisition of complex knowledge. Rather, the implementation of ongoing education to learn meta-cognitive strategies seems to be important if (re-)presented knowledge is to be integrated and used in a fruitful way (e.g., Hacker, Dunlosky, & Graesser, 1998).

The term *knowledge communication* comprises all processes that include the distribution of information and knowledge, the imparting of knowledge, the sharing and social construction of knowledge as well as knowledge-based cooperation (Reinmann-Rothmeier & Mandl, 2000). Since the different categories of knowledge management are highly intertwined, most problems and challenges of the communication of implicit knowledge have already been outlined for knowledge generation. But one aspect has to be stressed here: when knowledge is communicated the reliability of this knowledge is most important. Two phenomena have to be considered that might render knowledge unreliable. First, the motivation to communicate knowledge. On a personal level this motivation depends on the answer to the question: 'What do I gain or lose if I impart my knowledge?' On an organizational level the motivation to communicate knowledge might be reduced by the hidden agendas of individuals and groups within the organization (e.g., Baumard, 1999; Crozier & Friedberg, 1980; Williamson, 1993). In general, this problem is discussed within the realm of the 'principal–agent theory'[3] where hidden information and/or hidden actions may cause enormous problems and costs for organizations (e.g., Keser & Willinger, 2000; Lockwood,

[3] The principal–agent theory says that within an economy the division of labor leads to the differentiation that one person—the agent—performs work for another person—the principal. In this constellation a classical dilemma can be found: the principal never has complete information on the actions, intentions, and performance of the agent (e.g., situation of 'hidden information') and therefore the agent is encouraged not to fulfill his obligations toward the principal completely.

2000). Since this problem is not restricted to implicit knowledge it will not be outlined further.

Second, the ability to communicate implicit knowledge. For this area it is especially important to note that not every experienced person is an expert (e.g., Ericsson, Krampe & Tesch-Römer, 1993). If one combines this with results showing that experts have difficulties in verbalizing their knowledge (see Kirsner & Speelman, 1998) and that the explicit knowledge of experienced 'non-experts' contains more naive theories (see Herbig, 2001; Herbig & Büssing, 2002) that in turn can easily be verbalized, the danger that erroneous or problematic knowledge might be communicated becomes obvious. Moreover, on an organizational level Porac and Howard (1990) showed that the history of an organization may lead to the development of perception models that simplify cognition and thus promote erroneous knowledge in communication (see Baumard, 1999).

Another important problem of knowledge communication with regard to information- and communication systems (see knowledge generation) is implicit knowledge acquired and strengthened through concrete experience. Antonelli (1997) says that this technology is limited to the transfer of explicit (codifiable) knowledge; and Büssing and Herbig (1998) demonstrated for the domain of nursing care that the implementation of information and communication systems leads to a decrease of informal communication (at least regarding the content of interest; informal communication regarding the system itself may even increase, see Aydin & Rice, 1992), which in turn represses implicit knowledge. Therefore, the very (computerized) means to ensure knowledge communication might in themselves be hazardous to an important part of organizational knowledge (Johannessen et al., 2001).

The last process category of knowledge management is the *use of knowledge*. This category comprises the transformation of knowledge into decisions and actions whereby new knowledge can be generated. Here, the flexibility of implicit, experiential knowledge as proposed in work psychology (Carus et al., 1992; Martin, 1995) plays an important role. In order to gain this flexibility the use of knowledge in many different contexts seems to be necessary. A consequence of this assumption for organizational knowledge management is that members of the organization should get enough opportunities to use their knowledge in different areas and to experience the consequences of their actions. As implicit knowledge is not always adequate the risk of inadequate action in a real context can be too high. Therefore, in order to manage implicit knowledge, in its level of use, methods like the planning of games or simulations might be useful, as they allow people to benefit from the experience gained from the consequences of their use of knowledge without possible risks and costs for the organization.

To sum up (see Figure 7.3), implicit knowledge not only places high demands on the knowledge management in organizations but some of the common means of knowledge management (like information systems) also

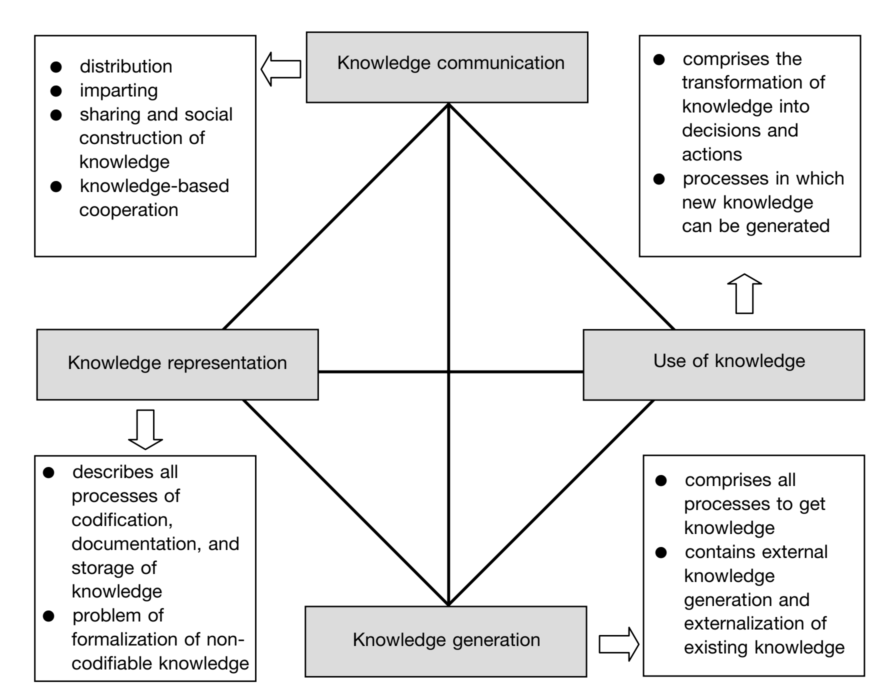


Figure 7.3 Process categories of knowledge management (data from Reinmann-Rothmeier & Mandl, 2000).

defy the essence of implicit knowledge itself. Moreover, in contrast to all other research approaches the practical approach of knowledge management seems to see no problem in the externalization of implicit knowledge.

IMPLICIT KNOWLEDGE: A REFINED DEFINITION AND AN INTEGRATIVE APPROACH

Summary of Research Findings and Refined Definition

As the presented approaches to the phenomenon of implicit knowledge show, there are not only slight disagreements on certain features of this type of knowledge but sometimes even contrasting opinions on what implicit knowledge is supposed to be as well. Table 7.1 summarizes the findings from different research directions. Because divergent findings or opinions can be found in each of these research directions, we will now try to name the most supported opinion.

Weighing the different findings from the different research areas, Büssing, Herbig, and Ewert (1999) came up with the following refined *definition of implicit knowledge*, which comprises various fundamental findings we consider to be essential for this type of knowledge:

Implicit knowledge contains declarative as well as procedural knowledge (Lewicki, 1986; Moss, 1995). It is acquired and strengthened by concrete and sensory experiences (Polanyi, 1966). Acquisition of implicit structures does not depend on attention or awareness for learning (Reber, 1997); moreover—as a direct consequence on this—its contents are not reflected and examined. One of the most prominent features of implicit knowledge is that it is not consciously perceived as guiding one's actions, that is, it works below a subjective threshold (Dienes & Berry, 1997). It also has a complex structure (Berry & Broadbent, 1988) and contains 'naive', sometimes wrong theories that can be examined and changed (Fischbein, 1994; Lee & Gelman, 1993; Sternberg, 1995) through explication (Gaines & Shaw, 1993).

Table 7.1 Summary of research findings.

	Cognitive psychology	Pedagogical psychology	Work psychology	Organizational psychology
Research methods	Laboratory experiments with artificial tasks	Intervention studies	Observation and interviews	Observation and interviews
Acquisition	Implicit learning	Early in the development	Apprenticeship experience in the real domain routinization	Apprenticeship experience in the real domain communication
Consciousness	Not conscious	/	Sometimes not conscious	Sometimes not conscious
Attention	Might not be necessary dependent on kind of task	/	/	/
Verbalizability	Not verbalizable	Mostly not verbalizable	Not verbalizable	Mostly verbalizable
Contents	Rules and contingencies for artifical tasks	Adequate or inadequate implicit naive theories in line with or contrasting explicit knowledge	Adequate holistic images including diffuse information, feelings, and sensations	Adequate (expert) knowledge for a certain domain
Complexity	Can be complex	/	Is complex	Is complex
Flexibility	Is not flexible	/	Is flexible	Is flexible

/ = no clear statement on this question.

In this definition one prominent feature of implicit knowledge—the lack of verbalizability—is not mentioned for two intertwined reasons. First, verbalizability is one of the most controversially described features of implicit knowledge (see Table 7.1), and, second, the concept has a close relation to the question of consciousness. As already described, one can make a difference between the 'no access' and the 'possible access' position. Since empirical evidence seems to support the 'possible access' position, it is assumed that at least part of implicit knowledge can be verbalized under certain conditions (e.g., if they are brought into the focus of awareness). Although the lack of verbalizability is—for the outlined reasons—not mentioned in the definition, it has to be stressed that the difficulty to express implicit knowledge is under natural conditions an important feature of this knowledge type.

An Integrative Approach

Looking at the summarized results in Table 7.1 it becomes evident that the different goals within each of the research directions led to differences in methods used in each approach, which in turn resulted in quite different assumptions on implicit knowledge. Roughly speaking, cognitive psychology deals with the fundamental structure and processes of implicit knowledge; work psychology tries to explore the special achievements of implicit knowledge at work; pedagogical and organizational psychology are mainly concerned with questions of knowledge imparting whereby organizational psychology has a special focus on managing implicit knowledge as a unique human resource. In order to gain a more complete insight into the phenomenon, it is necessary to integrate and expand these methods.

In cognitive psychology the common denominator of experimental tasks is the implicit learning of artificial rules that have no relation with knowledge from 'real' life, in order to ensure comparability of the knowledge bases of the test persons. This advantage turns into a disadvantage if one wants to compare explicit and implicit knowledge in a certain domain (e.g., at work). This limitation is quite problematic from the perspective of work psychology since empirical research regarding expertise (e.g., Sonnentag, 2000) and work experience (e.g., Benner & Tanner, 1987) indicates a close relation between explicitly learned, domain-specific knowledge and diffuse, somehow difficult-to-verbalize knowledge that is acquired implicitly during work.

On the other hand, typical methods for studying implicit knowledge in work psychology are interviews and observation whereby implicit knowledge is assumed if people are not able to give adequate explanations for their actions. Thus, with this approach it is difficult to understand the content and structure of implicit knowledge. Moreover, since 'impressive' demonstrations of implicit knowledge at work are most commonly

investigated, the question of problematic or even erroneous contents has not been studied.

Finally, the methods employed, mostly by pedagogical psychology, are intervention studies using different instructions in order to investigate their impact on knowledge and action (e.g., Mandl, De Corte, Bennett, & Friedrich, 1990). Here again, the content and structure of implicit knowledge is not at the center of research. However, differently from work psychology, problematic contents are investigated because they are a hindrance for learning.

Büssing et al. (1999) tried to integrate these different approaches and to overcome the mentioned problems from the perspective of work psychology. Three research objectives were followed. First, an investigation of the contents of implicit knowledge in a controlled experimental work setting. Second, the investigation of the relation between implicit and explicit professional knowledge. And, third, the comparison between successful and non-successful actors (for details on method development and validation see Büssing & Herbig, 2002; Büssing, Herbig & Ewert, 2002).

First, *investigation in a controlled setting*: to ensure a controlled setting, in which implicit knowledge could be used, the simulation of a critical situation (in the domain of nursing) was developed in cooperation with experts. This situation comprised features that should allow for the use of implicit knowledge as defined in work psychology (e.g., diffuse and sensory information). Students were carefully trained to act as patients, and the probands had to deal with the situations while recorded on video.

Second, *investigation of explicit and implicit knowledge*: to investigate explicit and implicit knowledge, tests for both knowledge types had to be developed. Explicit knowledge was investigated by means of a half-structured interview one to two weeks before the actual experimental session. The questions were ordered hierarchically from general to very specific to get further data on the accessibility of explicit knowledge.

For the explication of implicit knowledge the problem of verbalizability had to be taken into account. Therefore, important elements of the situation were used as a starting point for explication. That is, after dealing with the simulated situation a video-supported, cued recall of the situation takes place; and probands have to name elements from the situation that were important for their actions. These elements are then further explored by means of a repertory grid procedure that allows a (re-)construction of underlying (implicit) relations between the elements by dichotic constructs (see Kelly, 1969). The repertory grid technique is suitable for research into implicit knowledge based on experience because of its proximity to constructivism—Kelly's hypothesis that a person subjectively construes his or her world and that these constructs are not easy to access inter-individually. It can be used (e.g., after a critical incident) for describing in detail the knowledge relevant in the situation and to determine the action-guiding

constructs. These constructs are used according to Kelly (1955) to make forecasts about future events. The actual event shows whether these forecasts were correct or misleading, and offers the individual the possibility of revising or strengthening their constructs. This process based on the anticipation of events and the evaluation of the constructs can be regarded in our research as the process and use of experience. Every construct consists of a dichotic reference axis, which has both differentiating and integrating functions during the construction of the 'world'. The two functions 'differentiation' and 'integration' show large similarities to Polanyi's (1962) description of implicit knowledge—he also describes the two functions as fundamental processes of our construction of the world. In Polanyi's (1962) explanations the differentiating function appears as a rather explicit process, which presupposes conscious attention, while integration occurs without attention processes on an implicit level. This also justifies the suitability of the repertory grid method, since the method questions distinguishing constructs that are, according to Polanyi (1962), more easy to explicate, in order to highlight implicit knowledge in its integrated or integrating form.

In the third step of the explication process the repertory grids are visualized by means of correspondence analysis (e.g., Benzécri, 1992) so that validation of the contents of implicit knowledge as well as identification of problematic contents is possible.

In the next step, *comparison of successful and unsuccessful persons*: the clearly defined situation allowed the contrasting of successful and unsuccessful persons along criteria for the quality of action. Moreover, the relationship between implicit and explicit knowledge in those two groups of people could be investigated (e.g., via multidimensional scaling procedures: Young, 1985).

This integrative approach led to some interesting findings. First, a very high ecological validity showed that it is indeed possible to study complex working conditions in the laboratory. Moreover, it could be demonstrated that even complex, experiential implicit knowledge can be explicated (Büssing & Herbig, 2002; Büssing et al., 2002). Second, it was found that the quantity of explicit knowledge bore no relation to the quality of action meanwhile the quantity of implicit knowledge had an impact (Büssing et al., 2001). Third, it could be demonstrated that the implicit knowledge of successful persons was organized in a different way to the implicit knowledge of unsuccessful persons. Moreover, these differences in knowledge organization could be interpreted within the framework of experience-guided working; for example, unsuccessful persons organized their implicit knowledge along the time sequence of the situation while successful persons had a holistic organization where feelings had a diagnostical value for dealing with the situation (Herbig, Büssing, & Ewert, 2001). Fourth, direct comparison between explicit and implicit knowledge revealed that successful persons had a more complex explicit and a more flexible implicit knowledge than unsuccessful persons (Herbig, 2001; Herbig & Büssing, 2002).

In our future research the influence of the explication of implicit knowledge on performance will be investigated by a longitudinal design. Moreover, the effect of experience can be further studied by comparing experienced and unexperienced nurses (see Büssing et al., 2002a).

This short example of an integrative approach shows how methods and theories from different research areas can be combined in order to gain fuller understanding of the role of implicit knowledge at work. The concluding section will try to give some ideas on further research directions and the consequences of dealing with implicit knowledge at work and in organizations.

IMPLICATIONS FOR WORK AND ORGANIZATIONS, AND DIRECTIONS FOR FUTURE RESEARCH

In recent years implicit knowledge has been identified as a valuable human resource in organizations. It has been postulated that knowledge as an input resource will have greater impact than physical capital in the future (Drucker, 1993), and it has been estimated that up to 80% of knowledge in organizations is of implicit nature (Nonaka & Takeuchi, 1995). Besides the quantity of implicit knowledge, the quality of this knowledge creates sustainable competitive value (Johannessen et al., 2001) since it is bound to a person, difficult to verbalize, entrained in action, and linked to concrete contexts. Explicit knowledge can be more easily gained and transferred between organizations, but implicit knowledge is the unique resource of an organization. Therefore, implicit knowledge and its management are of great concern for work and organizations.

However, the distinction between the goals of organizations and the goals of scientific research into implicit learning has to be discussed. As outlined above research is interested in the special properties, advantages, and disadvantages of this knowledge type and its relation to other types of knowledge. Organizations, however, are most of the time quite naturally interested in obtaining and using this knowledge for the outlined reasons. This difference in goals can be specified by the difference in terminology used. Organizations and in particular companies want 'externalization', research is more interested in 'explication'; that is, as long as implicit knowledge is transferred within an organization the means are less relevant for organizations, whereas research needs data on the contents and structure of this knowledge. Although knowledge management has become an increasingly acknowledged necessity in companies many organizations still rely on the notion that the transfer of implicit knowledge simply happens (e.g., Pleskina, 2002). Others make more sophisticated attempts by giving their employees opportunities to externalize their knowledge. For example, the German weekly newspaper '*Die Zeit*' reported recently that a firm producing machine tools was not

prepared to let an employee retire, because this employee had developed his own highly successful manual grinding technique but was not able to explain it. After recognizing the importance of this 'implicit knowledge', admittedly quite late, the company's solution was to put an apprentice at the side of the employee for a year so that the apprentice could learn the technique (*Die Zeit*, 2002). This would preserve both implicit knowledge and competitive advantage for the firm, but only as long as the apprentice stayed there.

The attempts in organizations to transfer implicit knowledge might be categorized from 'it simply happens' or 'intervene if it might get lost' (see the machine tool example) to 'build an organizational frame in which implicit knowledge can be acquired and externalized'. We will take a closer look at the last category. Organizations that adopt this guideline mostly act within the maxim of 'experience promotion'. That is, they implicitly or explicitly adhere to the belief that experience builds up and shapes implicit knowledge, as Polanyi (1966) proposed. Experience promotion means the support and best possible promotion of acquisition, use, and exchange of experience (see Schulze, Witt, & Rose, 2002).

Thus, organizations trying to manage implicit knowledge set up basic conditions for this to happen. Three different approaches can be described that are used to promote experience and implicit knowledge. These approaches focus on (1) software, (2) hardware/tools, and (3) face-to-face communication in which there are some overlaps between the different approaches as the examples outlined below will show. Moreover, although the examples purely focus on one approach it has to be mentioned that most organizations follow a number of strategies with regard to knowledge management.

The software approach can be divided into the establishment of expert systems (see the section, 'Implicit knowledge in organizational psychology') and the development of information and communication systems. Information systems mostly have a focus on explicit knowledge while communication systems should enhance the probability of exchanging implicit knowledge. In this way they have a common denominator with the communication approach outlined below. The example describes electronic knowledge management (EKM), which specially claims an emphasis on implicit knowledge (Heinold, 2001). EKM is based on heterogeneous knowledge communities that should help their members to process the daily enormous amount of information more effectively and efficiently. Every participant of a heterogeneous knowledge community is therefore asked, as a first step, to take stock of his/her knowledge by answering the following questions: In which areas would I like to process information in order to obtain or acquire knowledge? What areas of knowledge are essential for my work? In what areas can I help the community members to get knowledge? As a second step, software solutions are developed that nominate those people who are knowledge offerors and those who are knowledge demanders. In the following steps, existing knowledge has to be prepared for 'online' use; for example, knowledge

about special orders that is collected in a file folder has to be computerized in a practical and customized way (in this example by scanning). It is important that computerization really presents the information in the way that the user needs it. Besides this more formal approach, 'informal' communication between members has to be facilitated by groupware that is integrated into the system. For example, the system should allow a knowledge demander to contact the appropriate knowledge offeror quickly or to discuss a certain problem with several people. In EKM it is stressed that intensive training of potential users is necessary for successful implementation.

The face-to-face communication approach stresses the importance of direct communication for the exchange and use of implicit knowledge as the following example from the Swiss Reinsurance Company (2000) will show. This company is one of the two biggest reinsurance companies worldwide and has more than 70 offices in 30 countries of the world and about 9,000 mostly highly qualified employees. It offers classical reinsurance cover, alternative risk transfer tools, and a spectrum of additional services for comprehensive management of capital and risk. Moreover, the company claims to attach key importance to its employees' know-how and learning, and thus to knowledge management. In the large German branch of the organization one of the important measures for knowledge management lies in an architecture that should enhance communication and knowledge transfer. Based on Winston Churchill's statement, 'First we form our buildings, then the buildings form us', an office concept was developed that should allow unimpeded exchange of information, give a communication-supporting environment, and allow teamwork as well as concentrated, solitary work. So, each team have the option to use different rooms such as concentration cells, team rooms, group offices, meeting zones and rooms, project rooms, and so-called technique islands (provided with fax, printer, and photocopying machines). The architectural structure is built in such a way that it can be completely reorganized within 24 hours. By locating archives and databases in commonly shared rooms, communication among employees about documented (explicit) and implicit knowledge should be strengthened. It is assumed that the processes of externalization and socialization are facilitated by these architectural measures according to the principle 'form follows flow'; that is, the flow of interaction and communication is the driving force behind architecture and work organization (Wittl, 2002).

A third and quite different approach is the measure to facilitate acquisition and use of implicit knowledge by changing hardware/tools in an experience-promoting way. Looking at the examples about CNC lathes (see the section, 'Implicit knowledge—the phenomenon'), it was shown that this technology is difficult to handle for employees since encasement of the machines hinders sensory perception of the working process. As sensory perception is an important part of experience-guided work, allowing the acquisition and use of implicit knowledge, one might assume that this technology could be a

hindrance to implicit knowledge use and acquisition. Based on extensive research into the special properties of experience-guided working and implicit knowledge in this area (for an overview see Martin, 1995; Schulze et al., 2002) several prototypical changes in CNC lathes were developed: for example, a resonance sensor that allows the worker to hear what is going on inside the machine; a so-called 'Rotoclear', that is, a windshield wiper for the machine window, which is normally opaque due to cooling solvent and flying-around cuttings, that allows the worker to see what is going on inside the machine; and a force feedback handwheel and override that allows the worker to feel the force or pressure necessary to bring the tool into contact with the material. This way of strengthening sensory perception and thereby implicit knowledge is also found in other areas. For example, force feedback joysticks in aviation or data gloves with haptic feedback for the control of medical robots. The difference between this approach and the two outlined above lies in another type of externalization of implicit knowledge. Once important parameters of an experience-guided working process are identified, this implicit knowledge is quasi-melted or coagulated into tools, which in turn should strengthen the acquisition and use of implicit knowledge by other workers. Subjective knowledge is 'objectified' and influences the further actions of people.

The presented examples demonstrate how one can try to accomplish externalization of implicit knowledge. Software-based communication systems and architectural support of communication among employees put in place the basic condition for an exchange of knowledge between people. Expert systems externalize the knowledge of acknowledged experts for the direct use and transfer of this knowledge. And new or adapted tools externalize implicit knowledge by objectifying it so that new knowledge might be generated by the use of these tools. However, by looking closely at these examples a problem emerges that might be formulated exaggeratedly as: organizations provide a framework for acquiring, using, changing, and transferring implicit knowledge and then they just hope for the best. Sometimes externalization might work but at other times it might be a problem, as the following arguments depict.

There are several pitfalls related to the management of implicit knowledge that should be considered. First, looking at the communication approaches, employees have to be ready to impart with their implicit knowledge; that is, their personal motivation should not be undermined by the evaluation that their labor market value is going to drop if they give their knowledge away. Moreover, the organizational climate and culture have to be such that a destructive demarcation of groups within the organization is not necessary (e.g., Cartwright, Cooper, & Earley, 2001; Dickson, Smith, Grojean, & Ehrhart, 2001). That is, a culture of organizational learning has to be implemented (e.g., Argote, Ingram, Levine, & Moreland, 2000; Argyris, 1999; Hayes & Allinson, 1998). Second, the problem of knowledge reliability might

be even more difficult to deal with. As outlined above, very experienced people or experts may have difficulty in expressing the knowledge they really use to solve a task or even worth, resulting in erroneous knowledge being imparted. By simply giving a framework for externalization, a closer look at the actual, imparted, implicit knowledge contents is normally not planned but might be necessary. How can organizations test the reliability of implicit knowledge?

Up to now this question has been restricted to the realm of expert systems and even there mistakes have happened (e.g., Kirsner & Speelman, 1998). As mentioned above, most organizations still rely on implicit knowledge transfer that simply 'happens' by socialization in groups (Nonaka & Takeuchi, 1995); that is, the transfer of implicit knowledge from one person to that of another person. Nonaka (1994) and Baumard (1999) explain knowledge creation in organizations as a spiraling process between explicit and implicit knowledge, working from the individual through the group through the organization into an inter-organizational dimension. However, as in most organizational theories, the question of the verbalizability necessary for explication as well as the question of reliability is not answered satisfactorily. For example, Baumard (1999, p. 24) merely states that: 'the conversion of tacit knowledge into explicit knowledge is realized daily in organizations'. Or, as Nonaka and Takeuchi (1995) explain, externalization is the process of articulation of implicit knowledge in explicit concepts. According to them, in this essential process, implicit knowledge takes on the form of metaphors, analogies, models, or hypotheses, which are often insufficient, illogical, or inadequate. These discrepancies between images and verbal expressions should then support collective reflection and interaction. There has never been an investigation into whether all relevant implicit knowledge can be phrased in pictures, which might indeed be a problem when considering the question of consciousness; nor has there been an explanation of the process of collective reflection and how this reflection might help in detecting inadequate knowledge.

Therefore, important research questions for the future seem to lie in utilization and adaptation of strict methods of cognitive psychology to investigate knowledge transfer processes in groups and organizations. Although research and companies do have different goals, the problem of knowledge reliability is important for both sides. For example, research and especially cognitive psychology can provide methods to explicate and visualize implicit knowledge (e.g., repertory grid, cognitive maps). This in turn might be used to transfer knowledge to other people. If these people are then able to solve a given problem with this knowledge one might assume that the explicated implicit knowledge is reliable. On the other hand, if the problem cannot be solved it is an indication that the knowledge is insufficient or erroneous, and thus unreliable. Especially in high-risk areas (like medicine or atomic plants) such a test is of crucial importance.

A related problem yet to be investigated concerns what happens with the special properties of implicit knowledge if it is externalized (i.e., made explicit). Results from work psychology show the importance of implicit experiential knowledge for dealing quickly and adequately with critical situations (e.g., Martin, 1995). Are such quick reactions still possible if the knowledge employed is explicit, or does this knowledge mode need more time for processing and thus hinders dealing with critical situations successfully? Moreover, Herbig (2001) showed that the reintegration of implicit knowledge that was externalized might cause problems for the persons 'receiving' this knowledge; for example, feelings as a reliable source of information in implicit knowledge tend to lose this positive function and to be disturbing if they reside in explicit knowledge. Research results from the CNC lathes example point to problems that might be similar: it took some time and practice before the workers were able to use the new or changed tools in a fruitful way (e.g., Carus, Schulze, & Ruppel, 1993); that is, new experience had to be amassed with use of the tools. In knowledge management this process is called internalization and is described as: people have to understand the experience of others (e.g., Nonaka & Takeuchi, 1995) in order to internalize knowledge. However, Polanyi (1962) stresses the importance of 'original' experience to integrate and use knowledge in a fruitful way. Here again, methods for closer investigation of the process have to be developed and intervention studies on the impact of certain learning ('internalization') conditions have to be conducted.

Put together, these questions demand and advocate a more basic research approach into implicit knowledge at work and in organizations (i.e., in the 'real world') in order to better understand the contents of this knowledge and the processes involved. An attempt at such an approach was outlined in the section 'An integrative approach'. Nevertheless, a more fundamental problem persists: a commonly acknowledged definition of implicit knowledge. As this chapter shows, the decision for or against a certain definitional property of implicit knowledge is quite arbitrary and depends largely on the viewpoint of the researcher. While our refined definition and integrative approach in the section, 'Summary of research findings and refined definition' comprises the important and essential properties of implicit knowledge we have to acknowledge that even the most commonly named features like 'not conscious' or 'not verbalizable' are handled in different ways. The situation becomes even more complicated by the different levels of analysis—individuals in artificial learning conditions, individuals in real situations, groups, and organizations.

Cognitive psychology with its aim of studying 'pure' implicit knowledge sidesteps one of the most difficult questions—the separation between implicit and explicit knowledge. If implicit knowledge is studied in the real domain, differentiation between what is implicit and what is explicit might not be completely possible. This notion even led some researchers to state that

implicit knowledge does not exist (e.g., Haider, 1991; Willingham & Preuss, 1995). However, we think that some of the confusion in defining implicit knowledge stems from the possible entanglement of implicit and explicit knowledge in reality and that an investigation of 'pure' implicit knowledge is not adequate on its own for research into work or organizational psychology (e.g., Mathews, 1997). Nevertheless, a scientific debate from different perspectives will be necessary to come to terms with the question of what 'really' defines implicit knowledge. With such a commonly acknowledged definition and the integration of different research methods, we would gain better insight into the advantages and possible disadvantages of implicit knowledge as well as starting points for dealing in an appropriate way with this type of knowledge.

ACKNOWLEDGMENT

Parts of this paper are based on research that was supported by the Deutsche Forschungsgemeinschaft (German Science Foundation) under BU/581-10-1, 10-2, 10-3 (principal investigator Univ.-Prof. Dr André Büssing) as part of the research group 'Knowledge and Behavior' (FO 204). We would like to thank Winfried Hacker (Technical University Dresden) and Sabine Sonnentag (Technical University Braunschweig) for their helpful comments on an earlier version of this paper.

REFERENCES

Ainley, P., & Rainbird, H. (eds) (1999). *Apprenticeship: Towards a New Paradigm of Learning*. London: Kogan Page.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, **89**, 369–406.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, **105**, 165–180.

Antonelli, C. (1997). New information technology and the knowledge-based economy. The Italian Evidence. *Review of Industrial Organization*, **12**, 593–607.

Argote, L., Ingram, P., Levine, J. M., & Moreland, R. L. (2000). Knowledge transfer in organizations. Learning from the experience of others. *Organizational Behavior and Human Decision Processes*, **82**, 1–8.

Argyris, C. (1999). *On Organizational Learning*. Oxford: Blackwell.

Aydin, C. E., & Rice, R. E. (1992). Bringing social worlds together. Computers as catalysts for new interactions in health care organizations. *Journal of Health and Social Behavior*, **33**, 168–185.

Baumard, P. (1999). *Tacit Knowledge in Organization*. Thousand Oaks, CA: Sage.

Benner P. (1984). *From Novice to Expert: Excellence and Power in Clinical Nursing Practice*. Reading, MA: Addison-Wesley.

Benner, P., & Tanner, C. (1987). How expert nurses use intuition. *American Journal of Nursing*, **1**, 23–31.

Benzécri, J. P. (1992). *Correspondence Analysis Handbook*. New York: Marcel Dekker.

Berry, D. C. (ed.) (1997). *How Implicit Is Implicit Learning?* Oxford: Oxford University Press.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit–explicit distinction. *British Journal of Psychology*, **79**, 251–272.

Böhle, F., & Milkau, B. (1988). *Vom Handrad zum Bildschirm. Eine Untersuchung zur sinnlichen Erfahrung im Arbeitsprozeß* [From Handwheel to Display. An Examination of Sensory Experience in the Working Process]. Frankfurt: Campus [in German].

Broadbent, D. E., FitzGerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, **77**, 33–50.

Browne, D. (1997). Putting knowledge to work. *Behavioral and Brain Sciences*, **20**, 353–354.

Büssing, A., & Herbig, B. (1998). The challenges of a care information system reflecting holistic nursing care. *Computers in Nursing*, **16**, 311–317.

Büssing, A., & Herbig, B. (2002). Tacit knowledge and experience in working. *Psychologische Beiträge* (in press).

Büssing, A., Herbig, B., & Ewert, T. (1999). *Implizites Wissen und erfahrungsgeleitetes Arbeitshandeln: Konzeptualisierung und Methodenentwicklung* [Tacit Knowledge and Experience-guided Working: Conceptualizing and Method Development, Report No. 48]. München: Technical University München, Chair of Psychology [in German].

Büssing, A., Herbig, B., & Ewert, T. (2001). Implizites und explizites Wissen—Einflüsse auf Handeln in kritischen Situationen [Implicit and explicit knowledge—influences on action in critical situations]. *Zeitschrift für Psychologie*, **209**, 174–200 [in German].

Büssing, A., Herbig, B., & Ewert, T. (2002). Implizites Wissen und erfahrungsgeleitetes Arbeitshandeln: Entwicklung einer Methode zur Explikation in der Krankenpflege [Implicit knowledge and experience-guided working: Development of a method for explication in nursing]. *Zeitschrift für Arbeits- und Organisationspsychologie*, **46**, 2–21 [in German].

Büssing, A., Herbig, B., & Latzel, A. (2002a). *Der Einfluss der Explikation impliziten Wissens auf die Handlungsgüte von Pflegekrften—eine experimentelle Untersuchung in der Krankenpflege*. [The Influence of the Explication of Implicit Knowledge on the Quality of Action—an Experimental Study in Nursing, Report No. 61]. München: Technical University München, Chair of Psychology [in German].

Büssing, A., Herbig, B., & Latzel, A. (2002b). Knowledge—experience—behavior. A new model to analyze their mutual relation. In D. Frey, H. Mandl, & L. von Rosenstiel (eds), *Knowledge and Behavior*. Seattle: Hogrefe/Huber Publishers (in press).

Campbell, M. (1990). Systematization of nursing and the promise of computers: A new phase in nurses' struggle for control of their practice? In E. Dimitz (ed.), *Computers in Hospital Care* (Proceedings, International Colloquium, 8–9 June). Wien: Österreichische Akademie der Wissenschaften.

Cartwright, S., Cooper, C., & Earley, C. (eds) (2001). *The International Handbook of Organizational Culture and Climate*. Chichester, UK: John Wiley & Sons.

Carus, U., Nogala, D., & Schulze, H. (1992). Experience-guided working: An undervalued resource for advanced manufacturing systems. In P. Brödner, & W. Karwowski (eds), *Ergonomics of Hybrid Automated Systems III* (Proceedings of the Third International Conference on Human Aspects of Advanced

Manufacturing and Hybrid Automation, Gelsenkirchen, Germany, 26–28 August, 1992, pp. 423–428). Amsterdam: Elsevier.

Carus, U., Schulze, H., & Ruppel, R. (1993). Der Drehprozess zum Greifen nahe [The lathing process within one's reach]. *Technische Rundschau*, **85**, 15–18 [in German].

Chan, C. (1992). Implicit cognitive processes. Theoretical issues and applications in computer systems design. Unpublished doctoral dissertation, University of Oxford.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, **36**, 387–395.

Cimino, J. J. (1999). Development of expertise in medical practice. In R. J. Sternberg, & J. A. Horvath (eds), *Tacit Knowledge in Professional Practice—Researchers and Practitioners Perspectives* (pp. 101–120). Mahwah, NJ: Lawrence Erlbaum.

Clement, J. (1994). Physical intuition and imaginistic simulation in expert problem solving. In D. Tirosh (ed.), *Implicit and Explicit Knowledge: An Educational Approach* (Human Development, Vol. 6, pp. 204–244). Norwood, NJ: Ablex.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (ed.), *Knowing, Learning and Instruction* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Crozier, M., & Friedberg, E. (1980). *Actors and Systems: The Politics of Collective Action*. Chicago: University of Chicago Press.

Darlington, K. (2000). *The Essence of Expert Systems*. Harlow, UK: Prentice Hall.

Dickson, M. W., Smith, D. B., Grojean, M. W., & Ehrhart, M. (2001). An organizational climate regarding ethics: The outcome of leader values and the practices that reflect them. *Leadership Quarterly*, **12**, 197–217.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, **4**, 3–23.

Dierkes, M., Antal, A. B., Child, J., & Nonaka, I. (eds). (2001). *Handbook of Organizational Learning and Knowledge*. Oxford: Oxford University Press.

Die Zeit (17 January 2002). Schürfen in der Datenhalde [Digging in the data stock] (p. 31). Hamburg: *Die Zeit* [in German].

Doane, S. M., Sohn, Y. W., & Schreiber, B. (1999). The role of processing strategies in the acquisition and transfer of cognitive skills. *Journal of Experimental Psychology—Human Perception and Performance*, **25**, 1390–1410.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.

Drucker, P. F. (1993). *Post-capitalist Society*. New York: Butterworth Heinemann.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, **100**, 363–406.

Ericsson, K. A., & Smith, J. (eds) (1991). *Toward a General Theory of Expertise*. Cambridge: Cambridge University Press.

Firlej, M., & Hellens, D. (1991). *Knowledge Elicitation: A Practical Handbook*. New York: Prentice Hall.

Fischbein, E. (1994). Tacit models. In D. Tirosh (ed.), *Implicit and Explicit Knowledge: An Educational Approach* (Human Development, Vol. 6, pp. 96–110). Norwood, NJ: Ablex.

Gaines, B. R., & Shaw, M. L. G. (1993). Knowledge acquisition tools based on personal construct psychology. *Knowledge Engineering Review*, **8**, 49–85.

Gazzard, S. (1994). *On Implicit Learning of Sequential and Conceptual Rules*. Sydney: University of Sydney, Department of Psychology.

Gelman, R. (1994). Constructivism and supporting environment. In D. Tirosh (ed.), *Implicit and Explicit Knowledge: An Educational Approach* (Human Development, Vol. 6, pp. 55–82). Norwood, NJ: Ablex.

Girotto, V., & Gonzales, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, **78**, 247–276.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, **53**, 5–26.

Greeno, J. G., Smith, D. R., & Moore, J. L. (1993). Transfer of situated learning. In D. K. Dettermann, & R. J. Sternberg (eds), *Transfer on trial: Intelligence, Cognition, and Instruction* (pp. 99–167). Norwood, NJ: Ablex.

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist*, **47**, 766–779.

Griffin, T., Schwartz, S., & Sofronoff, K. (1998). Implicit processes in medical diagnosis. In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, M. Anderson, & C. MacLeod (eds), *Implicit and Explicit Mental Processes* (pp. 329–341). Mahwah, NJ: Lawrence Erlbaum.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (eds) (1998). *Metacognition in Educational Theory and Practice*. Mahwah, NJ: Lawrence Erlbaum.

Hacker, W. (1992). *Expertenkönnen. Erkennen und Vermitteln* [Expert Skill. Detecting and Imparting]. Göttingen: Verlag für Angewandte Psychologie [in German].

Hacker, W. (1998). *Allgemeine Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten* [General Work Psychology. Psychological Regulation of Work Activities]. Bern: Hans Huber [in German].

Haider, H. (1991). *Die Unterscheidung zwischen explizitem und implizitem Wissen und Lernen* [The Differentiation between Explicit and Implicit Learning]. Hamburg: Universität der Bundeswehr, Institut für Kognitionswissenschaft [in German].

Hantano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (eds), *Child Development and Education in Japan* (pp. 262–272). San Francisco: Freeman.

Hartman, M., Knopman, D. S., & Nissen, M. J. (1989). Implicit learning of new verbal associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 1070–1082.

Hay, K. E., & Barab, S. A. (2001). Constructivism in practice: A comparison and contrast of apprenticeship and constructionist learning environments. *Journal of the Learning Sciences*, **10**, 281–322.

Hayes, J., & Allinson, C. W. (1998). Cognitive style and the theory and practice of individual and collective learning in organizations. *Human Relations*, **51**, 847–871.

Heinold, U. (2001). E-Knowledge-Management: Alles 'E'? [E-knowledge-management: Everything 'E'?]. *Wissensmanagement—Das Magazin für Führungskräfte*, **3**, 18–20 [in German].

Herbig, B. (2001). *Vergleichende Untersuchung von Struktur und Inhalt expliziten und impliziten Wissens im Arbeitskontext* [Comparative Study of Structure and Content of Explicit and Implicit Knowledge in the Work Context]. Aachen, Germany: Shaker [in German].

Herbig, B., & Büssing, A. (2002). Comparison of explicit and implicit knowledge in the work context. *Psychologische Beiträge* (in press).

Herbig, B., Büssing, A., & Ewert, T. (2001). The role of tacit knowledge in the work context of nursing. *Journal of Advanced Nursing*, **34**, 687–695.

Higgins, S., & Baumfield, V. (1998). A defense of teaching general thinking skills. *Journal of the Philosophy of Education*, **32**, 391–398.

Holyoak, K. J., & Spellman, B. A. (1993). Thinking. *Annual Review of Psychology*, **44**, 265–315.

Jackson, P. (1999). *Introduction to Expert Systems*. Harlow, UK: Addison-Wesley.

Johannessen, J.-A., & Hauan, A. (1994). Information, innovation and organizational learning. *Creativity and Innovation Management*, **3**(2).

Johannessen, J.-A., Olaisen, J., & Hauan, A. (1993). The challenge of innovation in a Norwegian shipyard: Facing the Russian market. *European Journal of Marketing*, **27**(3), 23–39.

Johannessen, J.-A., Olaisen, J., & Olsen, B. (2001). Mismanagement of tacit knowledge: The importance of tacit knowledge, the danger of information technology, and what to do about it. *International Journal of Information Management*, **21**, 3–20.

Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawings. *Cognition*, **34**, 57–83.

Kelly, G. A. (1955). *The Psychology of Personal Constructs*. New York: Norton.

Kelly, G. A. (1969). A mathematical approach to psychology. In B. Maher (ed.), *Clinical Psychology and Personality: The Selected Papers of George Kelly* (pp. 94–113). New York: John Wiley & Sons.

Keser, C., & Willinger, M. (2000). Principals' principles when agents' actions are hidden. *International Journal of Industrial Organization*, **18**, 163–185.

Kirsner, K., & Speelman, C. (1998). Introduction and overview. In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, M. Anderson, & C. MacLeod (eds), *Implicit and Explicit Mental Processes* (pp. 3–12). Mahwah, NJ: Lawrence Erlbaum.

Kirsner, K., Speelman, C., Maybery, M., O'Brien-Malone, A., Anderson, M., & MacLeod, C. (eds) (1998). *Implicit and Explicit Mental Processes*. Mahwah, NJ: Lawrence Erlbaum.

Kriegel, H.-P. (2000). Datenbanktechniken zur Unterstützung des Wissenserwerbs [Data bank techniques for the support of knowledge acquisition]. In H. Mandl, & G. Reinmann-Rothmeier (eds), *Wissensmanagement. Informationszuwachs—Wissensschwund? Die strategische Bedeutung des Wissensmanagements* [Knowledge management. Information increase—knowledge decrease? The strategical meaning of knowledge management] (pp. 47–71). München: Oldenbourg [in German].

Lant, T. K., & Shapira, Z. (eds). (2001). *Organizational Cognition: Computation and Interpretation*. Mahwah, NJ: Lawrence Erlbaum.

Lee, M. G., & Gelman, R. (1993). *Contributions of Developmental and Cognitive Psychology to Education* (Education and Technology Series monograph). Paris: United Nations Educational, Scientific, and Cultural Organization.

Lewicki, P. (1986). *Nonconscious Social Information Processing*. London: Academic Press.

Lewicki, P., Czyzewska, M., & Hill, T. (1997). Nonconscious information processing and personality. In D. C. Berry (ed.), *How Implicit Is Implicit Learning?* (pp. 48–72). Oxford: Oxford University Press.

Liebowitz, J. (ed.) (1998). *The Handbook of Applied Expert Systems*. Boca Raton, FL: CRC Press.

Lockwood, B. (2000). Production externalities and two-way distortion in principal–multi-agent problems. *Journal of Economic Theory*, **92**, 142–166.

Macchi, L. (1997). Pragmatically before ecologically valid tasks. *Behavioral and Brain Sciences*, **20**, 778–779.

MacLeod, C. (1998). Implicit perception: Perceptual processing without awareness. In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, M. Anderson, &

C. MacLeod (eds), *Implicit and Explicit Mental Processes* (pp. 57–78). Mahwah, NJ: Lawrence Erlbaum.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, **51**, 93–120.

Mandl, H., De Corte, E., Bennett, S. N., & Friedrich, H. F. (eds). (1990). *Learning and Instruction, Vol. 2.1—Social and Cognitive Aspects of Learning and Instruction*. Oxford: Pergamon Press.

Mandl, H., & Gerstenmaier, J. (eds) (2000). *Die Kluft zwischen Wissen und Handeln—Empirische und theoretische Lösungsansätze* [The gap between knowledge and action—Empirical and theoretical attempts for solution]. Göttingen, Germany: Hogrefe [in German].

Martin, H. (ed.) (1995). *CeA—Computergestützte erfahrungsgeleitete Arbeit* [Computer-supported Experience-guided Working]. Berlin: Springer [in German].

Mathews, R. C. (1997). Is research painting a biased picture of implicit learning? The dangers of methodological purity in scientific debate. *Psychonomic Bulletin & Review*, **4**, 38–42.

Menzies, T. (1998). Toward situated knowledge acquisition. *International Journal of Human-Computer Studies*, **49**, 867–893.

Moss, E. (1995). *The Grammar of Consciousness. An Exploration of Tacit Knowing*. London: Macmillan.

Myers, C., & Davis, K. (1993). Tacit skill and performance at work. *Applied Psychology*, **42**, 117–137.

Natsoulas, T. (1978). Consciousness. *American Psychologist*, **33**, 906–914.

Neal, A., & Hesketh, B. (1997). Future directions for implicit learning: Toward a clarification of issues asociated with knowledge representation and consciousness. *Psychonomic Bulletin & Review*, **4**, 73–78.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, **19**, 1–32.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, **5**, 14–37.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge Creating Company*. Oxford: Oxford University Press.

O'Brien-Malone, A., & Maybery, M. (1998). Implicit learning. In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, M. Anderson, & C. MacLeod (eds), *Implicit and Explicit Mental Processes* (pp. 37–55). Mahwah, NJ: Lawrence Erlbaum.

Olivera, F. (2000). Memory systems in organizations: An empirical investigation of mechanisms for knowledge collection, storage and access. *Journal of Management Studies*, **37**, 811–832.

Olson, D. R., & Campbell, R. (1994). Representation and mispresentation: On the beginning of symbolization in young children. In D. Tirosh (ed.), *Implicit and Explicit Knowledge: An Educational Approach* (Human Development, Vol. 6, pp. 83–95). Norwood, NJ: Ablex.

Pleskina, S. (2002). Explikation von Wissen—Mehrung von Erfahrungswissen im Unternehmen [Explication of knowledge—augmentation of tacit knowledge in companies]. Unpublished doctoral thesis, Technical University München, München [in German].

Polanyi, M. (1962). *Personal Knowledge: Toward a Post-critical Philosophy*. Chicago: University of Chicago Press.

Polanyi, M. (1966). *The Tacit Dimension*. New York: Doubleday.

Porac, J. F., & Howard, T. (1990). Taxonomic mental models in competitor definition. *Academy of Management Review*, **15**, 224–240 [in German].

Preussler, W. (1998). Strukturwissen als Voraussetzung für die Steuerung komplexer dynamischer Systeme [Structural knowledge as a precondition of controlling complex dynamic systems]. *Zeitschrift für experimentelle Psychologie*, **45**, 218–240 [in German].

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 88–94.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, **118**, 219–235.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge. An Essay on the Cognitive Unconscious*. New York: Oxford University Press.

Reber, A. S. (1997). Implicit ruminations. *Psychonomic Bulletin & Review*, **4**, 49–55.

Reinmann-Rothmeier, G., & Mandl, H. (2000). *Individuelles Wissensmanagement—Strategien für den persönlichen Umgang mit Information und Wissen am Arbeitsplatz* [Individual Knowledge Management—Strategies for a Personal Dealing with Information and Knowledge at the Workplace]. Göttingen, Germany: Hans Huber [in German].

Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird [Inert knowledge: When what is learned is not used]. *Psychologische Rundschau*, **47**, 78–92 [in German].

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, **31**, 115–121.

Rimann, M., Udris, I., & Weiss, V. (2000). Values in transition from apprenticeship to occupational work. *Swiss Journal of Psychology*, **59**, 291–302.

Roth, G., & Kleiner, A. (1998). Developing organizational memory through learning histories. *Organizational Dynamics*, **26**, 43–59.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental model. *Psychological Bulletin*, **100**, 349–363.

Ryle, G. (1993). *Aspects of Mind*. Oxford: Blackwell.

Sanders, A. F. (1988). *Michael Polanyi's Post-critical Epistemology. A Reconstruction of Some Aspects of "Tacit Knowledge"*. Amsterdam: Rodopi.

Sanderson, P. M. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 729–747.

Schacter, D. L., McAndrews, M. P., & Moscovitch, M. (1988). Access to consciousness: Dissociations between implicit and explicit knowledge in neuropsychological syndromes. In L. Weiskrantz (ed.), *Thought without Language* (pp. 242–278). London: Oxford University Press.

Schulze, H., Witt, H., & Rose, H. (2002). Erfahrungsförderlichkeit als ein Gestaltungsleitbild für Produktionstechnik und dessen Umsetzung [Experience promotion as a design model for production engineering and its realization]. In W. G. Weber, & T. Wehner (eds), *Erfahrungsorientierte Handlungsorganisation* [Experience-guided Organization of Action] (pp. 215–253). Zürich: VDF [in German].

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, **115**, 163–196.

Shanks, D. R., & St John, M. F. (1994). Characteristics of dissociable human memory systems. *Behavioral and Brain Sciences*, **17**, 367–447.

Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgements. *Organizational Behavior and Human Decision Processes*, **42**, 271–283.

Shimamura, A. P. (1989). Disorders of memory: The cognitive science perspective. In F. Boller, & J. Grafman (eds), *Handbook of Neuropsychology* (pp. 35–73). Amsterdam: Elsevier.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, **69**, 99–118.

Sloboda, J. (1991). Musical expertise. In K. A. Ericsson, & J. Smith (eds), *Toward a General Theory of Expertise* (pp. 153–171). Cambridge: Cambridge University Press.

Sonnentag, S. (2000). Expertise at work: Experience and excellent performance. In C. L. Cooper, & I. T. Robertson (eds), *International Review of Industrial and Organizational Psychology*, **15**, 223–264.

Speelman, C. (1998). Implicit expertise: Do we expect too much from our experts? In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, M. Anderson, & C. MacLeod (eds), *Implicit and Explicit Mental Processes* (pp. 135–147). Mahwah, NJ: Lawrence Erlbaum.

Sproull, L., & Kiesler, S. (1994). *Connections—New Ways of Working in Networked Organizations* (4th edn). Cambridge, MA: MIT Press.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, **127**, 161–188.

Sternberg, R. J. (1995). Theory and measurement of tacit knowledge as part of practical intelligence. *Zeitschrift für Psychologie*, **203**, 319–334.

Sternberg, R. J., & Horvath, J. A. (eds) (1999). *Tacit Knowledge in Professional Practice—Researcher and Practitioner Perspectives*. London: Lawrence Erlbaum.

Sveiby, K. E. (1997). *The New Organizational Wealth: Managing and Measuring Knowledge-based Assets*. San Francisco: Berrett-Koehler.

Swap, W., Leonard, D., Shields, M., & Abrams, L. (2001). Using mentoring and storytelling to transfer knowledge in the workplace. *Journal of Management Information Systems*, **18**, 95–114.

Swiss Reinsurance Company (2000). *Swiss Re 2000—Creating Value through Knowledge*. Zurich: Swiss Re Publishing.

Thurow, L. C. (1997). *The Future of Capitalism*. London: Nicholas Brealey.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, **23**, 727–741.

Wehner, T., & Waibel, M. (1996). Erfahrung als Bindeglied zwischen Handlungsfehleranalyse und Expertenforschung—Eine Studie am Schiffssimulator [Experience as link between action error analysis and research in expertise—a study at the ship simulator]. In J. Nitsch, & H. Allmer (eds), *Handeln im Sport—zwischen Rationalität und Intuition* [Acting in sports—between rationality and intuition] (pp. 115–139). Köln, Germany: bps-Verlag [in German].

Weinert, F. E., & Waldmann, M. R. (1988). Wissensentwicklung und Wissenserwerb [Knowledge development and knowledge acquisition]. In H. Mandl, & H. Spada (eds), *Wissenspsychologie* [Knowledge psychology] (pp. 161–199). Weinheim, Germany: Psychologie Verlags Union [in German].

Williamson, O. E. (1993). Calculativeness, trust and economic organization. *Journal of Law and Economics*, **30**, 131–145.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **15**, 1047–1060.

Willingham, D. B., & Preuss, L. (1995). The death of implicit memory. *Psyche*, **2**(15).

Wittl, A. (2002). Knowledge management. Strategies and team oriented forms of work organization at Swiss Re Germany AG (unpublished manuscript). München: Swiss Re Germany.

Wright, J. C., & Drinkwater, M. (1997). Rationality vs. accuracy of social judgement. *Social Cognition*, **15**, 245–273.

Young, F. W. (1985). Multidimensional scaling. In S. Kotz, & N. Johnson (eds), *Encyclopedia of Statistical Sciences*, Vol. 5 (pp. 649–658). New York: John Wiley & Sons.

INDEX

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson

International Review of Industrial and Organizational Psychology

CONTENTS OF PREVIOUS VOLUMES

VOLUME 17—2002

International Review of Industrial and Organizational Psychology 2003, Volume 18
Edited by Cary L. Cooper and Ivan T. Robertson

ISBN: 0-470-84703-4

VOLUME 16—2001

VOLUME 15—2000

VOLUME 14—1999

VOLUME 13—1998

VOLUME 12—1997

VOLUME 11—1996

VOLUME 10—1995

VOLUME 9—1994

VOLUME 8—1993

VOLUME 7—1992

VOLUME 6—1991

VOLUME 5—1990

VOLUME 4—1989

VOLUME 3—1988

VOLUME 2—1987

VOLUME 1—1986